

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-222000

(43)Date of publication of application : 09.08.2002

(51)Int.Cl.

G10L 19/12
H03M 7/36

(21)Application number : 2001-016870

(71)Applicant : SONY CORP

(22)Date of filing : 25.01.2001

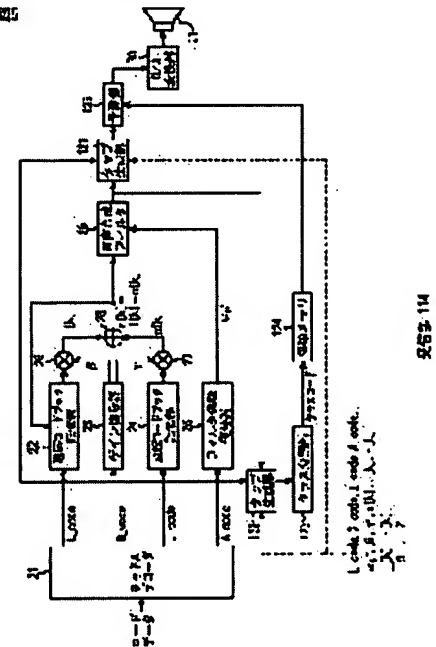
(72)Inventor : KONDO TETSUJIRO
KIMURA HIROTO
WATANABE TSUTOMU
HATTORI MASAOKI

(54) DEVICE AND METHOD FOR DATA PROCESSING, PROGRAM AND RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To obtain synthesized sound having good tone quality.

SOLUTION: A tap generating section 121 generates a prediction tap from the synthesized data of forty samples in the subframe of target data among the synthesized sound data decoded by the coded data of the voice by a CELP system and the synthesized sound data of forty samples whose starting point is located and delayed from a target subframe for the lag indicated by an L code arranged in the target subframe. Then, a predicting section 125 conducts a prescribed prediction computation by using the prediction tap and a tap coefficient stored in a coefficient memory 124 and high tone quality voice data are decoded.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the
examiner's decision of rejection or application
converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of
rejection][Date of requesting appeal against examiner's decision
of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-222000

(P2002-222000A)

(43) 公開日 平成14年8月9日(2002.8.9)

(51) Int.Cl.⁷

識別記号

F I

テーマコード(参考)

G 1 0 L 19/12

H 0 3 M 7/36

5 D 0 4 5

H 0 3 M 7/36

G 1 0 L 9/14

S 5 J 0 6 4

審査請求 未請求 請求項の数33 O L (全 34 頁)

(21) 出願番号 特願2001-16870(P2001-16870)

(71) 出願人 000002185

ソニー株式会社

東京都品川区北品川6丁目7番35号

(22) 出願日 平成13年1月25日(2001.1.25)

(72) 発明者 近藤 哲二郎

東京都品川区北品川6丁目7番35号 ソニー株式会社内

(72) 発明者 木村 裕人

東京都品川区北品川6丁目7番35号 ソニー株式会社内

(74) 代理人 100082131

弁理士 稲本 義雄

最終頁に続く

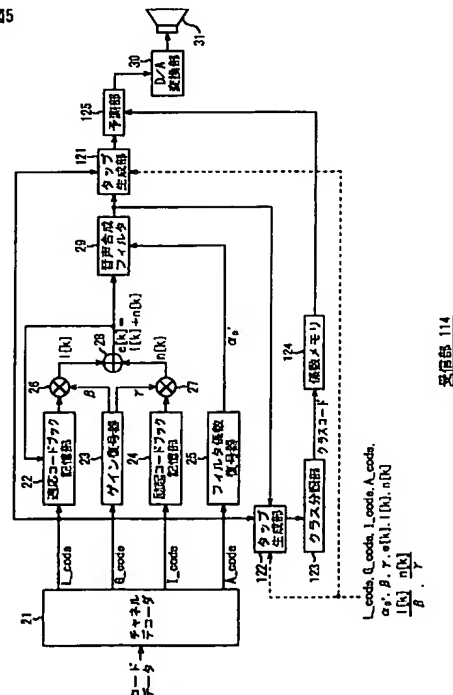
(54) 【発明の名称】 データ処理装置およびデータ処理方法、並びにプログラムおよび記録媒体

(57) 【要約】

【課題】 高音質の合成音等を得る。

【解決手段】 タップ生成部121は、CELP方式で符号化された音声の符号化データを復号した合成音データのうちの注目している注目データのサブフレームにおける40サンプルの合成音データと、注目サブフレームから、その注目サブフレームに配置されているLコードが表すラグだけ過去の位置を始点とする40サンプルの合成音データとから、予測タップを生成する。そして、予測部125は、その予測タップと、係数メモリ124に記憶されたタップ係数とを用いて、所定の予測演算を行うことにより、高音質の音声データを復号する。

図5



【特許請求の範囲】

【請求項1】 所定のデータと、周期を表す周期情報とを処理するデータ処理装置であって、前記所定のデータのうちの注目している注目データにつき、前記周期情報に応じて、前記所定のデータを抽出することにより、所定の処理に用いるタップを生成するタップ生成手段と、前記タップを用いて、前記注目データについて、所定の処理を行う処理手段とを備えることを特徴とするデータ処理装置。

【請求項2】 学習を行うことにより求められたタップ係数を取得するタップ係数取得手段をさらに備え、前記タップ生成手段は、前記タップ係数との所定の予測演算を行う予測タップを生成し、前記処理手段は、前記予測タップとタップ係数とを用いて、所定の予測演算を行うことにより、前記学習において教師として用いられた教師データに対応する予測値を求めることを特徴とする請求項1に記載のデータ処理装置。

【請求項3】 前記処理手段は、前記予測タップとタップ係数とを用いて、線形1次予測演算を行うことにより、前記予測値を求めることを特徴とする請求項2に記載のデータ処理装置。

【請求項4】 前記タップ生成手段は、前記注目データをクラス分けするクラス分類を行うのに用いるクラスタップを生成し、前記処理手段は、前記クラスタップに基づき、前記注目データについて、クラス分類を行うことを特徴とする請求項1に記載のデータ処理装置。

【請求項5】 前記タップ生成手段は、学習を行うことにより求められたタップ係数との所定の予測演算を行う予測タップを生成するとともに、前記注目データをクラス分けするクラス分類を行うのに用いられるクラスタップを生成し、前記処理手段は、前記クラスタップに基づき、前記注目データについて、クラス分類を行い、そのクラス分類の結果得られるクラスに対応する前記タップ係数と、前記予測タップとを用いて、所定の予測演算を行うことにより、前記学習において教師として用いられた教師データに対応する予測値を求めることを特徴とする請求項1に記載のデータ処理装置。

【請求項6】 前記所定のデータと周期情報は、音声を符号化した符号化データから得られるものであることを特徴とする請求項1に記載のデータ処理装置。

【請求項7】 前記符号化データは、音声を、CELP (Code Excited Liner Prediction coding)方式によって符号化したものであることを特徴とする請求項6に記載のデータ処理装置。

【請求項8】 前記周期情報は、CELP方式で規定されている長期予測ラグであることを特徴とする請求項7

に記載のデータ処理装置。

【請求項9】 前記所定のデータは、前記符号化データを復号した復号音声データであることを特徴とする請求項6に記載のデータ処理装置。

【請求項10】 前記所定のデータは、前記符号化データを音声データに復号するのに用いられる残差信号であることを特徴とする請求項6に記載のデータ処理装置。

【請求項11】 前記所定のデータは、時系列のデータであり、

10 前記タップ生成手段は、前記注目データから、前記周期情報に対応する時間だけ離れた位置の前記所定のデータを抽出することにより、前記タップを生成することを特徴とする請求項1に記載のデータ処理装置。

【請求項12】 前記タップ生成手段は、前記注目データから、前記周期情報に対応する時間だけ過去方向もしくは未来方向に離れた位置の前記所定のデータのうちの一方または両方を抽出することにより、前記タップを生成することを特徴とする請求項11に記載のデータ処理装置。

20 【請求項13】 前記所定のデータの波形の推移を判定する判定手段をさらに備え、前記タップ生成手段は、前記判定手段による判定結果に基づいて、前記周期情報に対応する時間だけ過去方向もしくは未来方向に離れた位置の前記所定のデータのうちの一方または両方を抽出することを特徴とする請求項12に記載のデータ処理装置。

【請求項14】 前記判定手段は、前記所定のデータのパワーに基づいて、その波形の推移を判定することを特徴とする請求項13に記載のデータ処理装置。

30 【請求項15】 所定のデータと、周期を表す周期情報とを処理するデータ処理方法であって、前記所定のデータのうちの注目している注目データにつき、前記周期情報に応じて、前記所定のデータを抽出することにより、所定の処理に用いるタップを生成するタップ生成ステップと、前記タップを用いて、前記注目データについて、所定の処理を行う処理ステップとを備えることを特徴とするデータ処理方法。

【請求項16】 所定のデータと、周期を表す周期情報とを、コンピュータに処理させるプログラムであって、前記所定のデータのうちの注目している注目データにつき、前記周期情報に応じて、前記所定のデータを抽出することにより、所定の処理に用いるタップを生成するタップ生成ステップと、前記タップを用いて、前記注目データについて、所定の処理を行う処理ステップとを備えることを特徴とするプログラム。

【請求項17】 所定のデータと、周期を表す周期情報とを、コンピュータに処理させるプログラムが記録されている記録媒体であって、

前記所定のデータのうちの注目している注目データにつき、前記周期情報に応じて、前記所定のデータを抽出することにより、所定の処理に用いるタップを生成するタップ生成ステップと、

前記タップを用いて、前記注目データについて、所定の処理を行う処理ステップとを備えるプログラムが記録されていることを特徴とする記録媒体。

【請求項 18】 所定のデータと、周期を表す周期情報とを処理するのに用いられる所定のタップ係数を学習するデータ処理装置であって、

学習の教師となる教師データから、前記所定のデータと周期情報を、学習の生徒となる生徒データとして生成する生徒データ生成手段と、

前記生徒データとしての所定のデータのうちの注目している注目データにつき、前記周期情報に応じて、前記所定のデータを抽出することにより、前記教師データを予測するのに用いる予測タップを生成する予測タップ生成手段と、

前記予測タップとタップ係数とを用いて、所定の予測演算を行うことにより得られる前記教師データの予測値の予測誤差が、統計的に最小になるように学習を行い、前記タップ係数を求める学習手段とを備えることを特徴とするデータ処理装置。

【請求項 19】 前記学習手段は、前記予測タップとタップ係数とを用いて、線形一次予測演算を行うことにより得られる前記教師データの予測値の予測誤差が、統計的に最小になるように学習を行うことを特徴とする請求項 18 に記載のデータ処理装置。

【請求項 20】 前記生徒データとしての所定のデータから、前記注目データをクラス分けするクラス分類を行うのに用いるクラスタップとするものを生成するクラスタップ生成手段と、

前記クラスタップに基づき、前記注目データについて、クラス分類を行うクラス分類手段とをさらに備え、

前記学習手段は、前記クラス分類手段によるクラス分類の結果得られるクラスごとに、前記タップ係数を求めることを特徴とする請求項 18 に記載のデータ処理装置。

【請求項 21】 前記クラスタップ生成手段は、前記注目データにつき、前記周期情報に応じて、前記所定のデータを抽出することにより、前記クラスタップを生成することを特徴とする請求項 20 に記載のデータ処理装置。

【請求項 22】 前記教師データは、音声データであり、前記所定のデータと周期情報は、前記教師データとしての音声データを符号化した符号化データから得られるものであることを特徴とする請求項 18 に記載のデータ処理装置。

【請求項 23】 前記符号化データは、音声データを、C E L P (Code Excited Liner Prediction coding)方式

によって符号化したものであることを特徴とする請求項 22 に記載のデータ処理装置。

【請求項 24】 前記周期情報は、C E L P 方式で規定されている長期予測ラグであることを特徴とする請求項 23 に記載のデータ処理装置。

【請求項 25】 前記所定のデータは、前記符号化データを復号した復号音声データであることを特徴とする請求項 22 に記載のデータ処理装置。

【請求項 26】 前記所定のデータは、前記符号化データを音声データに復号するのに用いられる残差信号であることを特徴とする請求項 22 に記載のデータ処理装置。

【請求項 27】 前記所定のデータは、時系列のデータであり、

前記予測タップ生成手段は、前記注目データから、前記周期情報に対応する時間だけ離れた位置の前記所定のデータを抽出することにより、前記予測タップを生成することを特徴とする請求項 18 に記載のデータ処理装置。

【請求項 28】 前記予測タップ生成手段は、前記注目データから、前記周期情報に対応する時間だけ過去方向もしくは未来方向に離れた位置の前記所定のデータのうちの一方または両方を抽出することにより、前記予測タップを生成することを特徴とする請求項 27 に記載のデータ処理装置。

【請求項 29】 前記所定のデータの波形の推移を判定する判定手段をさらに備え、

前記予測タップ生成手段は、前記判定手段による判定結果に基づいて、前記周期情報に対応する時間だけ過去方向もしくは未来方向に離れた位置の前記所定のデータのうちの一方または両方を抽出することを特徴とする請求項 28 に記載のデータ処理装置。

【請求項 30】 前記判定手段は、前記所定のデータのパワーに基づいて、その波形の推移を判定することを特徴とする請求項 29 に記載のデータ処理装置。

【請求項 31】 所定のデータと、周期を表す周期情報とを処理するのに用いられる所定のタップ係数を学習するデータ処理方法であって、

学習の教師となる教師データから、前記所定のデータと周期情報を、学習の生徒となる生徒データとして生成する生徒データ生成ステップと、

前記生徒データとしての所定のデータのうちの注目している注目データにつき、前記周期情報に応じて、前記所定のデータを抽出することにより、前記教師データを予測するのに用いる予測タップを生成する予測タップ生成ステップと、

前記予測タップとタップ係数とを用いて、所定の予測演算を行うことにより得られる前記教師データの予測値の予測誤差が、統計的に最小になるように学習を行い、前記タップ係数を求める学習ステップとを備えることを特徴とするデータ処理方法。

【請求項32】 所定のデータと、周期を表す周期情報とを処理するのに用いられる所定のタップ係数を学習するデータ処理を、コンピュータに行わせるプログラムであって、

学習の教師となる教師データから、前記所定のデータと周期情報を、学習の生徒となる生徒データとして生成する生徒データ生成ステップと、

前記生徒データとしての所定のデータのうちの注目している注目データにつき、前記周期情報に応じて、前記所定のデータを抽出することにより、前記教師データを予測するのに用いる予測タップを生成する予測タップ生成ステップと、

前記予測タップとタップ係数とを用いて、所定の予測演算を行うことにより得られる前記教師データの予測値の予測誤差が、統計的に最小になるように学習を行い、前記タップ係数を求める学習ステップとを備えることを特徴とするプログラム。

【請求項33】 所定のデータと、周期を表す周期情報とを処理するのに用いられる所定のタップ係数を学習するデータ処理を、コンピュータに行わせるプログラムが記録されている記録媒体であって、

学習の教師となる教師データから、前記所定のデータと周期情報を、学習の生徒となる生徒データとして生成する生徒データ生成ステップと、

前記生徒データとしての所定のデータのうちの注目している注目データにつき、前記周期情報に応じて、前記所定のデータを抽出することにより、前記教師データを予測するのに用いる予測タップを生成する予測タップ生成ステップと、

前記予測タップとタップ係数とを用いて、所定の予測演算を行うことにより得られる前記教師データの予測値の予測誤差が、統計的に最小になるように学習を行い、前記タップ係数を求める学習ステップとを備えるプログラムが記録されていることを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、データ処理装置およびデータ処理方法、並びにプログラムおよび記録媒体に関し、特に、例えば、CELP(Code Excited Linear Prediction coding)方式で符号化された音声、高音質の音声に復号することができるようにするデータ処理装置およびデータ処理方法、並びにプログラムおよび記録媒体に関する。

【0002】

【従来の技術】 図1および図2は、従来の携帯電話機の一例の構成を示している。

【0003】 この携帯電話機では、音声を、CELP方式により所定のコードに符号化して送信する送信処理と、他の携帯電話機から送信されてくるコードを受信して、音声に復号する受信処理とが行われるようになっており、図1は、送信処理を行う送信部を、図2は、受信処理を行う受信部を、それぞれ示している。

【0004】 図1に示した送信部では、ユーザが発話した音声、マイク（マイクロフォン）1に入力され、そこで、電気信号としての音声信号に変換され、A/D(Analog/Digital)変換部2に供給される。A/D変換部2は、マイク1からのアナログの音声信号を、例えば、8kHz等のサンプリング周波数でサンプリングすることにより、ディジタルの音声信号にA/D変換し、さらに、所定のビット数で量子化を行って、演算器3とLPC(Liner Prediction Coefficient)分析部4に供給する。

【0005】 LPC分析部4は、A/D変換部2からの音声信号を、例えば、160サンプル分の長さを1フレームとし、その1フレームを40サンプルごとのサブフレームに分割して、サブフレームごとにLPC分析し、P次の線形予測係数 $\alpha_1, \alpha_2, \dots, \alpha_P$ を求める。そして、LPC分析部4は、このP次の線形予測係数 α_p ($p=1, 2, \dots, P$)を要素とするベクトルを、音声の特徴ベクトルとして、ベクトル量子化部5に供給する。

【0006】 ベクトル量子化部5は、線形予測係数を要素とするコードベクトルとコードとを対応付けたコードブックを記憶しており、そのコードブックに基づいて、LPC分析部4からの特徴ベクトル α をベクトル量子化し、そのベクトル量子化の結果得られるコード（以下、適宜、Aコード(A_code)という)を、コード決定部15に供給する。

【0007】 さらに、ベクトル量子化部5は、Aコードに対応するコードベクトル α' を構成する要素となっている線形予測係数 $\alpha_1', \alpha_2', \dots, \alpha_P'$ を、音声合成フィルタ6に供給する。

【0008】 音声合成フィルタ6は、例えば、IIR(Infinite Impulse Response)型のディジタルフィルタで、ベクトル量子化部5からの線形予測係数 α_p' ($p=1, 2, \dots, P$)をIIRフィルタのタップ係数とするとともに、演算器14から供給される残差信号 e を入力信号として、音声合成を行う。

【0009】 即ち、LPC分析部4で行われるLPC分析は、現在時刻 n の音声信号（のサンプル値） s_n 、およびこれに隣接する過去のP個のサンプル値 $s_{n-1}, s_{n-2}, \dots, s_{n-P}$ に、式

$$s_n + \alpha_1 s_{n-1} + \alpha_2 s_{n-2} + \dots + \alpha_P s_{n-P} = e_n \quad \dots (1)$$

で示す線形1次結合が成立すると仮定し、現在時刻 n のサンプル値 s_n の予測値（線形予測値） s_n' を、過去の

P個の標本値 $s_{n-1}, s_{n-2}, \dots, s_{n-P}$ を用いて、式

7

8

$$s_n' = -(\alpha_1 s_{n-1} + \alpha_2 s_{n-2} + \dots + \alpha_p s_{n-p}) \quad \dots (2)$$

によって線形予測したときに、実際のサンプル値 s_n と線形予測値 s_n' との間の自乗誤差を最小にする線形予測係数 α_p を求めるものである。

【0010】ここで、式(1)において、 $\{e_n\}$ (\dots

$$s_n = e_n - (\alpha_1 s_{n-1} + \alpha_2 s_{n-2} + \dots + \alpha_p s_{n-p}) \quad \dots (3)$$

で表すことができ、これを、Z変換すると、次式が成立する。

$$S = E / (1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \dots + \alpha_p z^{-p}) \quad \dots (4)$$

但し、式(4)において、SとEは、式(3)における s_n と e_n のZ変換を、それぞれ表す。

$$e_n = s_n - s_n'$$

で表すことができ、実際のサンプル値 s_n と線形予測値 s_n' との間の残差信号と呼ばれる。

【0014】従って、式(4)から、線形予測係数 α_p をIIRフィルタのタップ係数とするとともに、残差信号 e_n をIIRフィルタの入力信号とすることにより、音声信号 s_n を求めることができる。

【0015】そこで、音声合成フィルタ6は、上述したように、ベクトル量子化部5からの線形予測係数 α_p' をタップ係数とするとともに、演算器14から供給される残差信号 e を入力信号として、式(4)を演算し、音声信号(合成音データ) s_s を求める。

【0016】なお、音声合成フィルタ6では、LPC分析部4によるLPC分析の結果得られる線形予測係数 α_p ではなく、そのベクトル量子化の結果得られるコードに対応するコードベクトルとしての線形予測係数 α_p' が用いられるため、音声合成フィルタ6が出力する合成音信号は、A/D変換部2が出力する音声信号とは、基本的に同一にはならない。

【0017】音声合成フィルタ6が出力する合成音データ s_s は、演算器3に供給される。演算器3は、音声合成フィルタ6からの合成音データ s_s から、A/D変換部2が出力する音声信号 s を減算し、その減算値を、自乗誤差演算部7に供給する。自乗誤差演算部7は、演算器3からの減算値の自乗和(第kサブフレームのサンプル値についての自乗和)を演算し、その結果得られる自乗誤差を、自乗誤差最小判定部8に供給する。

【0018】自乗誤差最小判定部8は、自乗誤差演算部7が出力する自乗誤差に対応付けて、長期予測ラグを表すコードとしてのLコード(L_code)、ゲインを表すコードとしてのGコード(G_code)、および符号語(励起コードブック)を表すコードとしてのIコード(I_code)を記憶しており、自乗誤差演算部7が出力する自乗誤差に対応するLコード、Gコード、およびIコードを出力する。Lコードは、適応コードブック記憶部9に、Gコー

$\dots, e_{n-1}, e_n, e_{n+1}, \dots$) は、平均値が0で、分散が所定値 σ^2 の互いに無相関な確率変数である。

【0011】式(1)から、サンプル値 s_n は、式

【0012】

【0013】ここで、式(1)および(2)から、 e_n は、式

\dots (5)

ドは、ゲイン復号器10に、Iコードは、励起コードブック記憶部11に、それぞれ供給される。さらに、Lコード、Gコード、およびIコードは、コード決定部15にも供給される。

【0019】適応コードブック記憶部9は、例えば7ビットのLコードと、所定の遅延時間(ラグ)とを対応付けた適応コードブックを記憶しており、演算器14から供給される残差信号 e を、自乗誤差最小判定部8から供給されるLコードに対応付けられた遅延時間(長期予測ラグ)だけ遅延して、演算器12に出力する。

【0020】ここで、適応コードブック記憶部9は、残差信号 e を、Lコードに対応する時間だけ遅延して出力することから、その出力信号は、その遅延時間を周期とする周期信号に近い信号となる。この信号は、線形予測係数を用いた音声合成において、主として、有声音の合成音を生成するための駆動信号となる。従って、Lコードは、概念的には、音声のピッチ周期を表す。なお、CELPの規格によれば、レコードは、20乃至146の範囲の整数値をとる。

【0021】ゲイン復号器10は、Gコードと、所定のゲイン β および γ とを対応付けたテーブルを記憶しており、自乗誤差最小判定部8から供給されるGコードに対応付けられたゲイン β および γ を出力する。ゲイン β と γ は、演算器12と13に、それぞれ供給される。ここで、ゲイン β は、長期フィルタ状態出力ゲインと呼ばれるものであり、また、ゲイン γ は、励起コードブックゲインと呼ばれるものである。

【0022】励起コードブック記憶部11は、例えば9ビットのIコードと、所定の励起信号とを対応付けた励起コードブックを記憶しており、自乗誤差最小判定部8から供給されるIコードに対応付けられた励起信号を、演算器13に出力する。

【0023】ここで、励起コードブックに記憶されている励起信号は、例えば、ホワイトノイズ等に近い信号で

あり、線形予測係数を用いた音声合成において、主として、無声音の合成音を生成するための駆動信号となる。

【0024】演算器12は、適応コードブック記憶部9の出力信号と、ゲイン復号器10が出力するゲイン β とを乗算し、その乗算値1を、演算器14に供給する。演算器13は、励起コードブック記憶部11の出力信号と、ゲイン復号器10が出力するゲイン γ とを乗算し、その乗算値nを、演算器14に供給する。演算器14は、演算器12からの乗算値1と、演算器13からの乗算値nとを加算し、その加算値を、残差信号eとして、

音声合成フィルタ6と適応コードブック記憶部9に供給する。

【0025】音声合成フィルタ6では、以上のようにして、演算器14から供給される残差信号eを入力信号が、ベクトル量子化部5から供給される線形予測係数 α_p をタップ係数とするIIRフィルタでフィルタリングされ、その結果得られる合成音データが、演算器3に供給される。そして、演算器3および自乗誤差演算部7において、上述の場合と同様の処理が行われ、その結果得られる自乗誤差が、自乗誤差最小判定部8に供給される。

【0026】自乗誤差最小判定部8は、自乗誤差演算部7からの自乗誤差が最小（極小）になったかどうかを判定する。そして、自乗誤差最小判定部8は、自乗誤差が最小になっていないと判定した場合、上述のように、その自乗誤差に対応するLコード、Gコード、およびIコードを出力し、以下、同様の処理が繰り返される。

【0027】一方、自乗誤差最小判定部8は、自乗誤差が最小になったと判定した場合、確定信号を、コード決定部15に出力する。コード決定部15は、ベクトル量子化部5から供給されるAコードをラッチするとともに、自乗誤差最小判定部8から供給されるLコード、Gコード、およびIコードを順次ラッチするようになっており、自乗誤差最小判定部8から確定信号を受信すると、そのときラッチしているAコード、Lコード、Gコード、およびIコードを、チャンネルエンコーダ16に供給する。チャンネルエンコーダ16は、コード決定部15からのAコード、Lコード、Gコード、およびIコードを多重化し、コードデータとして出力する。このコードデータは、伝送路を介して送信される。

【0028】以上から、コードデータは、復号に用いられる情報であるAコード、Lコード、Gコード、およびIコードを、サブフレーム単位ごとに有する符号化データとなっている。

【0029】なお、ここでは、Aコード、Lコード、Gコード、およびIコードは、サブフレームごとに求められるものとしているが、例えば、Aコードについては、フレームごとに求められる場合があり、この場合、そのフレームを構成する4つのサブフレームの復号には、同一のAコードが用いられる。但し、この場合でも、その

1フレームを構成する4つのサブフレームそれぞれが、同一のAコードを有していると見ることができ、そのように考えることによって、コードデータは、復号に用いられる情報であるAコード、Lコード、Gコード、およびIコードを、サブフレーム単位ごとに有する符号化データとなっているとみなすことができる。

【0030】ここで、図1（後述する図2、図5、図9、図11、図16、図18、および図21においても同様）では、各変数に、[k]が付され、配列変数とされている。このkは、サブフレーム数を表すが、明細書中では、その記述は、適宜省略する。

【0031】次に、以上のようにして、他の携帯電話機の送信部から送信されてくるコードデータは、図2に示した受信部のチャンネルデコーダ21で受信される。チャンネルデコーダ21は、コードデータから、Lコード、Gコード、Iコード、Aコードを分離し、それぞれを、適応コードブック記憶部22、ゲイン復号器23、励起コードブック記憶部24、フィルタ係数復号器25に供給する。

【0032】適応コードブック記憶部22、ゲイン復号器23、励起コードブック記憶部24、演算器26乃至28は、図1の適応コードブック記憶部9、ゲイン復号器10、励起コードブック記憶部11、演算器12乃至14とそれぞれ同様に構成されるもので、図1で説明した場合と同様の処理が行われることにより、Lコード、Gコード、およびIコードが、残差信号eに復号される。この残差信号eは、音声合成フィルタ29に対して、入力信号として与えられる。

【0033】フィルタ係数復号器25は、図1のベクトル量子化部5が記憶しているのと同じコードブックを記憶しており、Aコードを、線形予測係数 α_p に復号し、音声合成フィルタ29に供給する。

【0034】音声合成フィルタ29は、図1の音声合成フィルタ6と同様に構成されており、フィルタ係数復号器25からの線形予測係数 α_p をタップ係数とするとともに、演算器28から供給される残差信号eを入力信号として、式(4)を演算し、これにより、図1の自乗誤差最小判定部8において自乗誤差が最小と判定されたときの合成音データを生成する。この合成音データは、D/A(Digital/Analog)変換部30に供給される。D/A変換部30は、音声合成フィルタ29からの合成音データを、デジタル信号からアナログ信号にD/A変換し、スピーカ31に供給して出力させる。

【0035】なお、コードデータにおいて、Aコードが、サブフレーム単位でなく、フレーム単位で配置されている場合、図2の受信部では、フレームを構成する4つのサブフレームすべての復号に、そのフレームに配置されたAコードに対応する線形予測係数を用いることができる他、各サブフレームについて、隣接するフレームのAコードに対応する線形予測係数を用いて補間を行

い、その補間の結果得られる線形予測係数を、各サブフレームの復号に用いることが可能である。

【0036】

【発明が解決しようとする課題】以上のように、携帯電話機の送信部では、受信部の音声合成フィルタ 29 に与えられる入力信号としての残差信号と線形予測係数がコード化されて送信されてくるため、受信部では、そのコードが、残差信号と線形予測係数に復号される。しかしながら、この復号された残差信号や線形予測係数（以下、適宜、それぞれを、復号残差信号または復号線形予測係数という）には、量子化誤差等の誤差が含まれるため、音声を用いて分析して得られる残差信号と線形予測係数には一致しない。

【0037】このため、受信部の音声合成フィルタ 29 が出力する合成音データは、歪み等を有する、音質の劣化したものとなる。

【0038】本発明は、このような状況に鑑みてなされたものであり、高音質の合成音等を得ることができるようにするものである。

【0039】

【課題を解決するための手段】本発明の第 1 のデータ処理装置は、所定のデータのうちの注目している注目データにつき、周期情報に応じて、所定のデータを抽出することにより、所定の処理に用いるタップを生成するタップ生成手段と、タップを用いて、注目データについて、所定の処理を行う処理手段とを備えることを特徴とする。

【0040】本発明の第 1 のデータ処理方法は、所定のデータのうちの注目している注目データにつき、周期情報に応じて、所定のデータを抽出することにより、所定の処理に用いるタップを生成するタップ生成ステップと、タップを用いて、注目データについて、所定の処理を行う処理ステップとを備えることを特徴とする。

【0041】本発明の第 1 のプログラムは、所定のデータのうちの注目している注目データにつき、周期情報に応じて、所定のデータを抽出することにより、所定の処理に用いるタップを生成するタップ生成ステップと、タップを用いて、注目データについて、所定の処理を行う処理ステップとを備えることを特徴とする。

【0042】本発明の第 1 の記録媒体は、所定のデータのうちの注目している注目データにつき、周期情報に応じて、所定のデータを抽出することにより、所定の処理に用いるタップを生成するタップ生成ステップと、タップを用いて、注目データについて、所定の処理を行う処理ステップとを備えるプログラムが記録されていることを特徴とする。

【0043】本発明の第 2 のデータ処理装置は、学習の教師となる教師データから、所定のデータと周期情報を、学習の生徒となる生徒データとして生成する生徒データ生成手段と、生徒データとしての所定のデータのう

ちの注目している注目データにつき、周期情報に応じて、所定のデータを抽出することにより、教師データを予測するのに用いる予測タップを生成する予測タップ生成手段と、予測タップとタップ係数とを用いて、所定の予測演算を行うことにより得られる教師データの予測値の予測誤差が、統計的に最小になるように学習を行い、タップ係数を求める学習手段とを備えることを特徴とする。

【0044】本発明の第 2 のデータ処理方法は、学習の教師となる教師データから、所定のデータと周期情報を、学習の生徒となる生徒データとして生成する生徒データ生成ステップと、生徒データとしての所定のデータのうちの注目している注目データにつき、周期情報に応じて、所定のデータを抽出することにより、教師データを予測するのに用いる予測タップを生成する予測タップ生成ステップと、予測タップとタップ係数とを用いて、所定の予測演算を行うことにより得られる教師データの予測値の予測誤差が、統計的に最小になるように学習を行い、タップ係数を求める学習ステップとを備えることを特徴とする。

【0045】本発明の第 2 のプログラムは、学習の教師となる教師データから、所定のデータと周期情報を、学習の生徒となる生徒データとして生成する生徒データ生成ステップと、生徒データとしての所定のデータのうちの注目している注目データにつき、周期情報に応じて、所定のデータを抽出することにより、教師データを予測するのに用いる予測タップを生成する予測タップ生成ステップと、予測タップとタップ係数とを用いて、所定の予測演算を行うことにより得られる教師データの予測値の予測誤差が、統計的に最小になるように学習を行い、タップ係数を求める学習ステップとを備えることを特徴とする。

【0046】本発明の第 2 の記録媒体は、学習の教師となる教師データから、所定のデータと周期情報を、学習の生徒となる生徒データとして生成する生徒データ生成ステップと、生徒データとしての所定のデータのうちの注目している注目データにつき、周期情報に応じて、所定のデータを抽出することにより、教師データを予測するのに用いる予測タップを生成する予測タップ生成ステップと、予測タップとタップ係数とを用いて、所定の予測演算を行うことにより得られる教師データの予測値の予測誤差が、統計的に最小になるように学習を行い、タップ係数を求める学習ステップとを備えるプログラムが記録されていることを特徴とする。

【0047】本発明の第 1 のデータ処理装置およびデータ処理方法、並びにプログラムおよび記録媒体においては、所定のデータのうちの注目している注目データにつき、周期情報に応じて、所定のデータを抽出することにより、所定の処理に用いるタップが生成され、そのタップを用いて、注目データについて、所定の処理が行われ

る。

【0048】本発明の第2のデータ処理装置およびデータ処理方法、並びにプログラムおよび記録媒体においては、学習の教師となる教師データから、所定のデータと周期情報が、学習の生徒となる生徒データとして生成される。そして、生徒データとしての所定のデータのうちの注目している注目データにつき、周期情報に応じて、所定のデータを抽出することにより、教師データを予測するのに用いる予測タップが生成され、予測タップとタップ係数とを用いて、所定の予測演算を行うことにより得られる教師データの予測値の予測誤差が、統計的に最小になるように学習が行われ、タップ係数が求められる。

【0049】

【発明の実施の形態】図3は、本発明を適用した伝送システム（システムとは、複数の装置が論理的に集合した物をいい、各構成の装置が同一筐体中にあるか否かは問わない）の一実施の形態の構成を示している。

【0050】この伝送システムでは、携帯電話機101₁と101₂が、基地局102₁と102₂それぞれとの間で、無線による送受信を行うとともに、基地局102₁と102₂それぞれが、交換局103との間で送受信を行うことにより、最終的には、携帯電話機101₁と101₂との間において、基地局102₁および102₂、並びに交換局103を介して、音声の送受信を行うことができるようになっていく。なお、基地局102₁と102₂は、同一の基地局であっても良いし、異なる基地局であっても良い。

【0051】ここで、以下、特に区別する必要がない限り、携帯電話機101₁と101₂を、携帯電話機101と記述する。

【0052】次に、図4は、図3の携帯電話機101の構成例を示している。

【0053】この携帯電話機101では、CELP方式によって、音声の送受信が行われるようになっていく。

【0054】即ち、アンテナ111は、基地局102₁または102₂からの電波を受信し、その受信信号を、変復調部112に供給するとともに、変復調部112か

$$E[y] = w_1 x_1 + w_2 x_2 + \dots$$

【0060】式(6)を一般化するために、タップ係数 w_j の集合でなる行列 W 、生徒データ x_{ij} の集合でなる行列 X 、および予測値 $E[y_j]$ の集合でなる行列 Y' を、

【数1】

$$\begin{aligned} & \dots (6) \\ X &= \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1J} \\ x_{21} & x_{22} & \dots & x_{2J} \\ \dots & \dots & \dots & \dots \\ x_{I1} & x_{I2} & \dots & x_{IJ} \end{bmatrix} \\ W &= \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_J \end{bmatrix}, Y' = \begin{bmatrix} E[y_1] \\ E[y_2] \\ \dots \\ E[y_J] \end{bmatrix} \end{aligned}$$

で定義すると、次のような観測方程式が成立する。

50 【0061】

らの信号を、電波で、基地局102₁または102₂に送信する。変復調部112は、アンテナ111からの信号を復調し、その結果得られる、図1で説明したようなコードデータを、受信部114に供給する。また、変復調部112は、送信部113から供給される、図1で説明したようなコードデータを変調し、その結果得られる変調信号を、アンテナ111に供給する。送信部113は、図1に示した送信部と同様に構成され、そこに入力されるユーザの音声を、CELP方式によって、コードデータに符号化して、変復調部112に供給する。受信部114は、変復調部112からのコードデータを受信してCELP方式により復号し、さらに、高音質の音声を復号して出力する。

【0055】即ち、受信部114では、例えば、クラス分類適応処理を利用して、CELP方式で復号された合成音が、さらに、真の高音質の音声（の予測値）に復号される。

【0056】ここで、クラス分類適応処理は、クラス分類処理と適応処理とからなり、クラス分類処理によって、データを、その性質に基づいてクラス分けし、各クラスごとに適応処理を施すものであり、適応処理は、以下のような手法のものである。

【0057】即ち、適応処理では、例えば、合成音と、所定のタップ係数との線形結合により、高音質の音声の予測値が求められる。

【0058】具体的には、例えば、いま、高音質の音声（のサンプル値）を教師データとするとともに、その高音質の音声を、CELP方式によって、Lコード、Gコード、Iコード、およびAコードに符号化し、それらのコードを、図2に示した受信部で復号することにより得られる合成音を生徒データとして、教師データである高音質の音声 y の予測値 $E[y]$ を、幾つかの合成音（のサンプル値） x_1, x_2, \dots の集合と、所定のタップ係数 w_1, w_2, \dots の線形結合により規定される線形1次結合モデルにより求めることを考える。この場合、予測値 $E[y]$ は、次式で表すことができる。

【0059】

$$XW=Y'$$

ここで、行列Xの成分 x_{ij} は、i件目の生徒データの集合(i件目の教師データ y_i の予測に用いる生徒データの集合)の中のj番目の生徒データを意味し、行列Wの成分 w_j は、生徒データの集合の中のj番目の生徒データとの積が演算されるタップ係数を表す。また、 y_i は、i件目の教師データを表し、従って、 $E[y_i]$ は、i件目の教師データの予測値を表す。なお、式

(6)の左辺における y は、行列Yの成分 y_i のサフィックスiを省略したものであり、また、式(6)の右辺における x_1, x_2, \dots も、行列Xの成分 x_{ij} のサフィックスiを省略したものである。

【0062】そして、この観測方程式に最小自乗法を適用して、真の高音質の音声 y に近い予測値 $E[y]$ を求

$$XW=Y+E$$

【0064】この場合、元の高音質の音声 y に近い予測値 $E[y]$ を求めるためのタップ係数 w_j は、自乗誤差

$$\sum_{i=1}^I e_i^2$$

を最小にすることで求めることができる。

$$\theta_1 \frac{\partial \theta_1}{\partial w_j} + \theta_2 \frac{\partial \theta_2}{\partial w_j} + \dots + \theta_I \frac{\partial \theta_I}{\partial w_j} = 0 \quad (j=1, 2, \dots, J)$$

・・・(9)

【0067】そこで、まず、式(8)を、タップ係数 w_j で微分することにより、次式が成立する。

【0068】

【数5】

$$\frac{\partial \theta_i}{\partial w_1} = x_{i1}, \quad \frac{\partial \theta_i}{\partial w_2} = x_{i2}, \quad \dots, \quad \frac{\partial \theta_i}{\partial w_J} = x_{iJ}, \quad (i=1, 2, \dots, I)$$

・・・(10)

【0069】式(9)および(10)より、式(11)が得られる。

【0070】

【数6】

$$\sum_{i=1}^I \theta_i x_{i1} = 0, \quad \sum_{i=1}^I \theta_i x_{i2} = 0, \quad \dots, \quad \sum_{i=1}^I \theta_i x_{iJ} = 0$$

・・・(11)

【0071】さらに、式(8)の残差方程式における生徒データ x_{ij} 、タップ係数 w_j 、教師データ y_i 、および誤差 e_i の関係を考慮すると、式(11)から、次のような正規方程式を得ることができる。

【0072】

【数7】

・・・(7)

めることを考える。この場合、教師データとなる真の高音質の音声 y の集合でなる行列Y、および高音質の音声 y に対する予測値 $E[y]$ の残差 e の集合でなる行列Eを、

【数2】

$$E = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \dots \\ \theta_I \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_I \end{bmatrix}$$

で定義すると、式(7)から、次のような残差方程式が成立する。

【0063】

・・・(8)

【0065】従って、上述の自乗誤差をタップ係数 w_j で微分したものが0になる場合、即ち、次式を満たすタップ係数 w_j が、元の高音質の音声 y に近い予測値 $E[y]$ を求めるため最適値ということになる。

【0066】

【数4】

20

30

$$\begin{cases} \left(\sum_{i=1}^I x_{i1} x_{i1} \right) w_1 + \left(\sum_{i=1}^I x_{i1} x_{i2} \right) w_2 + \dots + \left(\sum_{i=1}^I x_{i1} x_{iJ} \right) w_J = \left(\sum_{i=1}^I x_{i1} y_i \right) \\ \left(\sum_{i=1}^I x_{i2} x_{i1} \right) w_1 + \left(\sum_{i=1}^I x_{i2} x_{i2} \right) w_2 + \dots + \left(\sum_{i=1}^I x_{i2} x_{iJ} \right) w_J = \left(\sum_{i=1}^I x_{i2} y_i \right) \\ \dots \\ \left(\sum_{i=1}^I x_{iJ} x_{i1} \right) w_1 + \left(\sum_{i=1}^I x_{iJ} x_{i2} \right) w_2 + \dots + \left(\sum_{i=1}^I x_{iJ} x_{iJ} \right) w_J = \left(\sum_{i=1}^I x_{iJ} y_i \right) \end{cases}$$

・・・(12)

【0073】なお、式(12)に示した正規方程式は、行列(共分散行列)Aおよびベクトル v を、

【数8】

$$A = \begin{pmatrix} \sum_{i=1}^I x_{i1}x_{i1} & \sum_{i=1}^I x_{i1}x_{i2} & \cdots & \sum_{i=1}^I x_{i1}x_{iJ} \\ \sum_{i=1}^I x_{i2}x_{i1} & \sum_{i=1}^I x_{i2}x_{i2} & \cdots & \sum_{i=1}^I x_{i2}x_{iJ} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^I x_{iJ}x_{i1} & \sum_{i=1}^I x_{iJ}x_{i2} & \cdots & \sum_{i=1}^I x_{iJ}x_{iJ} \end{pmatrix} \quad (17)$$

$$v = \begin{pmatrix} \sum_{i=1}^I x_{i1}y_i \\ \sum_{i=1}^I x_{i2}y_i \\ \vdots \\ \sum_{i=1}^I x_{iJ}y_i \end{pmatrix}$$

$$AW = v$$

で表すことができる。

【0074】式(12)における各正規方程式は、生徒データ x_{ij} および教師データ y_i のセットを、ある程度の数だけ用意することで、求めるべきタップ係数 w_j の数 J と同じ数だけたてることができ、従って、式(13)を、ベクトル W について解くことで(但し、式(13)を解くには、式(13)における行列 A が正則である必要がある)、最適なタップ係数(ここでは、自乗誤差を最小にするタップ係数) w_j を求めることができる。なお、式(13)を解くに当たっては、例えば、掃き出し法(Gauss-Jordanの消去法)などを用いることが可能である。

【0075】以上のようにして、最適なタップ係数 w_j を求めておき、さらに、そのタップ係数 w_j を用い、式(6)により、真の高音質の音声 y に近い予測値 $E[y]$ を求めるのが適応処理である。

【0076】なお、例えば、教師データとして、高いサンプリング周波数でサンプリングした音声信号、または多ビットを割り当てた音声信号を用いるとともに、生徒データとして、その教師データとしての音声信号を間引いたり、低ビットで再量子化した音声信号をC E L P方式により符号化し、その符号化結果を復号して得られる合成音を用いた場合、タップ係数としては、高いサンプリング周波数でサンプリングした音声信号、または多ビットを割り当てた音声信号を生成するのに、予測誤差が、統計的に最小となる高音質の音声を得られることになる。従って、この場合、より高音質の合成音を得ることが可能となる。

【0077】図4の受信部114では、以上のようなクラス分類適応処理により、コードデータを復号して得ら

18

で定義するとともに、ベクトル W を、数1で示したように定義すると、式

・・・(13)

れる合成音を、さらに、高音質の音声に復号するようになっている。

【0078】即ち、図5は、図4の受信部114の第1の構成例を示している。なお、図中、図2における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。

【0079】タップ生成部121と122には、音声合成フィルタ29が出力する、サブフレームごとの合成音データと、チャンネルデコーダ21が出力する、サブフレームごとのLコード、Gコード、Iコード、AコードのうちのLコードとが供給されるようになっている。タップ生成部121と122は、そこに供給される合成音データから、高音質の音声の予測値を予測するのに用いる予測タップとするものと、クラス分類に用いるクラスタップとするものを、Lコードに基づいて、それぞれ抽出する。予測タップは、予測部125に供給され、クラスタップは、クラス分類部123に供給される。

【0080】クラス分類部123は、タップ生成部122から供給されるクラスタップに基づいて、クラス分類を行い、そのクラス分類結果としてのクラスコードを、係数メモリ124に供給する。

【0081】ここで、クラス分類部123におけるクラス分類の方法としては、例えば、KビットADRC(Adaptive Dynamic Range Coding)処理を利用した方法等がある。

【0082】ここで、KビットADRC処理においては、例えば、クラスタップを構成するデータの最大値MAXと最小値MINが検出され、 $DR=MAX-MIN$ を、集合の局所的なダイナミックレンジとし、このダイナミックレンジDRに基づいて、クラスタップを構成する各データがKビットに再量子化される。即ち、クラスタップを構成する各デー

タから、最小値MINが減算され、その減算値が $DR/2^4$ で除算（量子化）される。そして、以上のようにして得られる、クラスタップを構成する各データのKビットの値を、所定の順番で並べたビット列が、ADRCコードとして出力される。

【0083】このようなKビットADRC処理を、クラス分類に利用する場合には、例えば、そのKビットADRC処理の結果得られるADRCコードを、クラスコードとすることが可能である。

【0084】なお、クラス分類は、その他、例えば、クラスタップを、それを構成する各データを要素とするベクトルと見なして、そのベクトルとしてのクラスタップをベクトル量子化すること等によって行うことも可能である。

【0085】係数メモリ124は、後述する図9の学習装置において学習処理が行われることにより得られる、クラスごとのタップ係数を記憶しており、クラス分類部123が出力するクラスコードに対応するアドレスに記憶されているタップ係数を、予測部125に供給する。

【0086】予測部125は、タップ生成部121が出力する予測タップと、係数メモリ124が出力するタップ係数とを取得し、その予測タップとタップ係数とを用いて、式（6）に示した線形予測演算を行う。これにより、予測部125は、注目している注目サブフレームについて、高音質の音声（の予測値）を求めて、D/A変換部30に供給する。

【0087】次に、図6のフローチャートを参照して、図5の受信部114の処理について説明する。

【0088】チャンネルデコーダ21は、そこに供給されるコードデータから、Lコード、Gコード、Iコード、Aコードを分離し、それぞれを、適応コードブック記憶部22、ゲイン復号器23、励起コードブック記憶部24、フィルタ係数復号器25に供給する。さらに、Lコードは、タップ生成部121および122にも供給される。

【0089】そして、適応コードブック記憶部22、ゲイン復号器23、励起コードブック記憶部24、演算器26乃至28は、図2における場合と同様の処理を行い、これにより、Lコード、Gコード、およびIコードが、残差信号eに復号される。この残差信号は、音声合成フィルタ29に供給される。

【0090】さらに、フィルタ係数復号器25は、図2で説明したように、そこに供給されるAコードを、線形予測係数に復号し、音声合成フィルタ29に供給する。音声合成フィルタ29は、演算器28からの残差信号と、フィルタ係数復号器25からの線形予測係数を用いて音声合成を行い、その結果得られる合成音を、タップ生成部121と122に供給する。

【0091】タップ生成部121は、音声合成フィルタ29が順次出力する合成音のサブフレームを、順次、注

目サブフレームとし、ステップS1において、その注目サブフレームの合成音データを抽出するとともに、その注目サブフレームから見て時間的に過去方向または未来方向の合成音データを、そこに供給されるLコードに基づいて抽出することにより、予測タップを生成し、予測部125に供給する。さらに、ステップS1では、タップ生成部122が、例えば、やはり、注目サブフレームの合成音データを抽出するとともに、その注目サブフレームから見て時間的に過去方向または未来方向の合成音データを、そこに供給されるLコードに基づいて抽出することにより、クラスタップを生成し、クラス分類部123に供給する。

【0092】そして、ステップS2に進み、クラス分類部123は、タップ生成部122から供給されるクラスタップに基づいて、クラス分類を行い、その結果得られるクラスコードを、係数メモリ124に供給して、ステップS3に進む。

【0093】ステップS3では、係数メモリ124は、クラス分類部123から供給されるクラスコードに対応するアドレスから、タップ係数を読み出し、予測部125に供給する。

【0094】そして、ステップS4に進み、予測部125は、係数メモリ124が出力するタップ係数を取得し、そのタップ係数と、タップ生成部121からの予測タップとを用いて、式（6）に示した積和演算を行い、注目サブフレームの高音質の音声データ（の予測値）を得る。

【0095】なお、ステップS1乃至S4の処理は、注目サブフレームの合成音データのサンプル値それぞれを、注目データとして行われる。即ち、サブフレームの合成音データは、前述したことから、40サンプルで構成されるから、その40サンプルの合成音データそれぞれについて、ステップS1乃至S4の処理が行われる。

【0096】以上のようにして得られた高音質の音声データは、予測部125から、D/A変換部30を介して、スピーカ31に供給され、これにより、スピーカ31からは、高音質の音声出力される。

【0097】ステップS4の処理後は、ステップS5に進み、まだ、注目サブフレームとして処理すべきサブフレームがあるかどうか判定され、あると判定された場合、ステップS1に戻り、次に注目サブフレームとすべきサブフレームを、新たに注目サブフレームとして、以下、同様の処理を繰り返す。また、ステップS5において、注目サブフレームとして処理すべきサブフレームがないと判定された場合、処理を終了する。

【0098】次に、図7および図8を参照して、図5のタップ生成部121での予測タップの生成方法について説明する。

【0099】タップ生成部121は、例えば、図7に示すように、注目サブフレームにおける40サンプルの合

成音データを抽出するとともに、注目サブフレームから、その注目サブフレームに配置されているＬコードが表すラグだけ過去の位置を始点とする４０サンプルの合成音データ（以下、適宜、ラグ対応の過去データという）を抽出し、注目データについての予測タップとする。

【０１００】あるいは、タップ生成部１２１は、例えば、図８に示すように、注目サブフレームの４０サンプルの合成音データを抽出するとともに、Ｌコードが表すラグだけ過去の位置が注目サブフレーム内の合成音データ（例えば、注目データ等）の位置となるＬコードが配置された、注目サブフレームから見て未来方向のサブフレームに配置された４０サンプルの合成音データ（以下、適宜、ラグ対応の未来データという）を抽出し、注目データについての予測タップとする。

【０１０１】また、タップ生成部１２１は、例えば、注目サブフレームの合成音データ、ラグ対応の過去データ、およびラグ対応の未来データを抽出し、注目データについての予測タップとする。

【０１０２】ここで、注目データを、クラス分類適応処理によって予測するにあたり、注目サブフレームの合成音データの他、注目サブフレーム以外のサブフレームの合成音データを、予測タップとして用いることにより、より高音質の音声を得ることができると考えられる。そして、この場合、予測タップは、例えば、単純に、注目サブフレームの合成音データの他、注目サブフレームの直前と直後のサブフレームの合成音データによって構成することが考えられる。

【０１０３】しかしながら、このように、予測タップを、単純に、注目サブフレームの合成音データ、および注目サブフレームの直前と直後のサブフレームの合成音データによって構成する場合には、予測タップの構成の仕方に、合成音データの波形特性がほとんど考慮されていないことから、その分、高音質化に影響が生じると考えられる。

【０１０４】そこで、タップ生成部１２１は、上述のように、予測タップとする合成音データを、Ｌコードに基づいて抽出するようになっている。

【０１０５】即ち、サブフレームに配置されているＬコードが表すラグ（長期予測ラグ）は、注目データ部分の合成音の波形が、過去のどの時点における合成音の波形に類似しているかを表すので、注目データの部分の波形と、ラグ対応の過去データやラグ対応の未来データの部分の波形とは、大きな相関性を有する。

【０１０６】従って、予測タップを、注目サブフレームの合成音データと、その合成音データに対して大きな相関を有するラグ対応の過去データもしくはラグ対応の未来データのうちの一方または両方とによって構成することにより、さらに高音質の音声を得ることが可能となる。

【０１０７】ここで、図５のタップ生成部１２２においても、例えば、タップ生成部１２１における場合と同様にして、注目サブフレームの合成音データと、ラグ対応の過去データもしくはラグ対応の未来データのうちの一方または両方とからクラスタップを生成することが可能であり、図５の実施の形態では、そのようになっている。

【０１０８】なお、予測タップやクラスタップの構成パターンは、上述したパターンのものに限定されるものではない。即ち、予測タップやクラスタップには、注目サブフレームのすべての合成音データを含める他、１サンプルおき等の合成音データだけを含めるようにしたり、注目サブフレームに配置されているＬコードが表すラグだけ過去の位置のサブフレームから、そのサブフレームに配置されているＬコードが表すラグだけ過去の位置のサブフレームの合成音データ等を含めるようにすることも可能である。

【０１０９】また、上述の場合には、クラスタップと予測タップを同一構成にするようにしたが、クラスタップと予測タップは、異なる構成とすることができる。

【０１１０】さらに、上述の場合には、Ｌコードが表すラグだけ過去の位置が注目サブフレーム内の合成音データ（例えば、注目データ等）の位置となるＬコードが配置された、注目サブフレームから見て未来方向のサブフレームに配置された４０サンプルの合成音データを、ラグ対応の未来データとして、予測タップに含めるようにしたが、ラグ対応の未来データとしては、その他、例えば、次のような合成音データを用いるようにすることも可能である。

【０１１１】即ち、ＣＥＬＰ方式において符号化データに含まれるＬコードは、上述したように、そのＬコードが配置されているサブフレームの合成音データの波形と類似する過去の合成音データの位置を表すが、符号化データには、そのような過去の類似波形の位置を表すＬコードの他、未来の類似波形の位置を表すＬコード（以下、適宜、未来Ｌコードという）を含めるようにすることが可能である。この場合、注目データについてのラグ対応の未来データとしては、注目サブフレームに配置されている未来Ｌコードが表すラグだけ未来の位置にある合成音データを始点とする１以上のサンプルを用いることが可能である。

【０１１２】次に、図９は、図５の係数メモリ１２４に記憶させるタップ係数の学習処理を行う学習装置の一実施の形態の構成例を示している。

【０１１３】マイク２０１乃至コード決定部２１５は、図１のマイク１乃至コード決定部１５とそれぞれ同様に構成される。マイク１には、学習用の音声信号が入力されるようになっており、従って、マイク２０１乃至コード決定部２１５では、その学習用の音声信号に対して、図１における場合と同様の処理が施される。

【0114】但し、コード決定部215は、Lコード、Gコード、Iコード、およびAコードのうち、本実施の形態において予測タップやクラスタップを構成する合成音データを抽出するのに用いられるLコードだけを出力するようになっている。

【0115】そして、タップ生成部131と132には、自乗誤差最小判定部208において自乗誤差が最小になったと判定されたときの音声合成フィルタ206が出力する合成音データが供給される。さらに、タップ生成部131と132には、コード決定部215が、自乗誤差最小判定部208から確定信号を受信したときに出力するLコードも供給される。また、正規方程式加算回路134には、A/D変換部202が出力する音声データが、教師データとして供給される。

【0116】タップ生成部131は、音声合成フィルタ206が出力する合成音データから、コード決定部215が出力するLコードに基づいて、図5のタップ生成部121における場合と同一の予測タップを生成し、生徒データとして、正規方程式加算回路134に供給する。

【0117】タップ生成部132も、音声合成フィルタ206が出力する合成音データから、コード決定部215が出力するLコードに基づいて、図5のタップ生成部122における場合と同一のクラスタップを生成し、クラス分類部133に供給する。

【0118】クラス分類部133は、タップ生成部132からのクラスタップに基づいて、図5のクラス分類部123における場合と同様のクラス分類を行い、その結果得られるクラスコードを、正規方程式加算回路134に供給する。

【0119】正規方程式加算回路134は、A/D変換部202からの音声データを、教師データとして受信するとともに、タップ生成部131からの予測タップを、生徒データとして受信し、その教師データおよび生徒データを対象として、クラス分類部133からのクラスコードごとに足し込みを行う。

【0120】即ち、正規方程式加算回路134は、クラス分類部133から供給されるクラスコードに対応するクラスごとに、予測タップ（生徒データ）を用い、式（13）の行列Aにおける各コンポーネントとなっている、生徒データどうしの乗算（ $x_{i,n} \cdot x_{j,n}$ ）と、サメーション（ Σ ）に相当する演算を行う。

【0121】さらに、正規方程式加算回路134は、やはり、クラス分類部133から供給されるクラスコードに対応するクラスごとに、生徒データおよび教師データを用い、式（13）のベクトルvにおける各コンポーネントとなっている、生徒データと教師データの乗算（ $x_{i,n} \cdot y_i$ ）と、サメーション（ Σ ）に相当する演算を行う。

【0122】正規方程式加算回路134は、以上の足し込みを、そこに供給される学習用の音声データのサブフ

レームすべてを注目サブフレームとし、かつその注目サブフレームの音声データすべてを注目データとして行い、これにより、各クラスについて、式（13）に示した正規方程式をたてる。

【0123】タップ係数決定回路135は、正規方程式加算回路134においてクラスごとに生成された正規方程式を解くことにより、クラスごとに、タップ係数を求め、係数メモリ136の、各クラスに対応するアドレスに供給する。

【0124】なお、学習用の音声信号として用意した音声信号によっては、正規方程式加算回路134において、タップ係数を求めるのに必要な数の正規方程式が得られないクラスが生じる場合があり得るが、タップ係数決定回路135は、そのようなクラスについては、例えば、デフォルトのタップ係数を出力する。

【0125】係数メモリ136は、タップ係数決定回路135から供給されるクラスごとのタップ係数を、そのクラスに対応するアドレスに記憶する。

【0126】次に、図10のフローチャートを参照して、図9の学習装置で行われる、高音質の音声を復号するためのタップ係数を求める学習処理について説明する。

【0127】学習装置には、学習用の音声信号が供給され、ステップS11では、その学習用の音声信号から、教師データと生徒データが生成される。

【0128】即ち、学習用の音声信号は、マイク201に入力され、マイク201乃至コード決定部215は、図1のマイク1乃至コード決定部15における場合とそれぞれ同様の処理を行う。

【0129】その結果、A/D変換部202で得られるデジタル信号の音声データは、教師データとして、正規方程式加算回路134に供給される。また、自乗誤差最小判定部208において自乗誤差が最小になったと判定されたときに、音声合成フィルタ206が出力する合成音データは、生徒データとして、タップ生成部131と132に供給される。さらに、自乗誤差最小判定部208において自乗誤差が最小になったと判定されたときに、コード決定部215が出力するLコードも、生徒データとして、タップ生成部131と132に供給される。

【0130】その後、ステップS12に進み、タップ生成部131は、音声合成フィルタ206から生徒データとして供給される合成音のサブフレームを注目サブフレームとし、さらに、その注目サブフレームの合成音データを、順次、注目データとして、各注目データについて、音声合成フィルタ206からの合成音データを用い、コード決定部215からのLコードに基づいて、図5のタップ生成部121における場合と同様にして、予測タップを生成し、正規方程式加算回路134に供給する。さらに、ステップS12では、タップ生成部132

が、やはり、合成音データを用い、Lコードに基づいて、図5のタップ生成部122における場合と同様にし、クラスタップを生成し、クラス分類部133に供給する。

【0131】ステップS12の処理後は、ステップS13に進み、クラス分類部133が、タップ生成部132からのクラスタップに基づいて、クラス分類を行い、その結果得られるクラスコードを、正規方程式加算回路134に供給する。

【0132】そして、ステップS14に進み、正規方程式加算回路134は、A/D変換器202からの教師データとしての高音質の音声データである学習用の音声データのうちの注目データに対応するもの、およびタップ生成部132からの生徒データとしての予測タップを対象として、式(13)の行列Aとベクトルvの、上述したような足し込みを、クラス分類部133からの注目データについてのクラスコードごとに行い、ステップS15に進む。

【0133】ステップS15では、まだ、注目サブフレームとして処理すべきサブフレームがあるかどうか判定される。ステップS15において、まだ、注目サブフレームとして処理すべきサブフレームがあると判定された場合、ステップS11に戻り、次のサブフレームを新たに注目サブフレームとして、以下、同様の処理が繰り返される。

【0134】また、ステップS15において、注目サブフレームとして処理すべきサブフレームがないと判定された場合、ステップS16に進み、タップ係数決定回路135は、正規方程式加算回路134で各クラスごとに生成された正規方程式を解くことにより、各クラスごとに、タップ係数を求め、係数メモリ136の、各クラスに対応するアドレスに供給して記憶させ、処理を終了する。

【0135】以上のようにして、係数メモリ136に記憶された各クラスごとのタップ係数が、図5の係数メモリ124に記憶されている。

【0136】以上のように、図5の係数メモリ124に記憶されたタップ係数は、線形予測演算を行うことにより得られる高音質の音声予測値の予測誤差(自乗誤差)が、統計的に最小になるように学習を行うことにより求められたものであるから、図5の予測部125が出力する音声は、高音質のものとなる。

【0137】なお、例えば、図5および図9の実施の形態では、予測タップやクラスタップを、音声合成フィルタ206が出力する合成音データから構成するようにしたが、予測タップやクラスタップは、図5および図9において点線で示すように、Iコードや、Lコード、Gコード、Aコード、Aコードから得られる線形予測係数 α 、Gコードから得られるゲイン β 、 γ 、その他の、Lコード、Gコード、Iコード、またはAコードから得ら

れる情報(例えば、残差信号eや、残差信号eを得るための l 、 n 、さらには、 l/β 、 n/γ など)のうちの1以上を含めて構成することが可能である。また、CELP方式では、符号化データとしてのコードデータに、リスト補間ビットやフレームエネルギー等が含まれる場合があるが、この場合、予測タップやクラスタップは、ソフト補間ビットやフレームエネルギー等を含めて構成することも可能である。

【0138】次に、図11は、図4の受信部114の第2の構成例を示している。なお、図中、図5における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。即ち、図11の受信部114は、タップ生成部121と122に替えて、タップ生成部301と302がそれぞれ設けられている他は、図5における場合と同様に構成されている。

【0139】図5の実施の形態においては、タップ生成部121や122において(図9のタップ生成部131と132においても同様)、予測タップやクラスタップが、注目サブフレームにおける40サンプルの合成音データの他、ラグ対応の過去データもしくはラグ対応の未来データのうちの一方または両方から構成されるが、ラグ対応の過去データだけ、ラグ対応の未来データだけ、またはその両方のうちのいずれを予測タップやクラスタップに含めるかを、特に制御するようにはなっていないことから、いずれを含めるかを、あらかじめ決めておき、固定しておく必要がある。

【0140】しかしながら、注目サブフレームを含むフレーム(以下、適宜、注目フレームという)等が、例えば、発話の開始時等に相当する場合には、図12(A)に示すように、注目フレームよりも過去のフレーム等は、無音の状態(雑音だけが存在するのと等しい状態)となっていると考えられる。同様に、注目フレームが、例えば、発話の終了時等に相当する場合には、図12(B)に示すように、注目フレームよりも未来のフレーム等は、無音の状態となっていると考えられる。そして、このような無音部分については、予測タップやクラスタップに含めても、音質の向上にほとんど寄与せず、むしろ、最悪の場合には、音質の向上を妨げるおそれがある。

【0141】一方、注目フレームが、発話の開始時や終了時等以外の定常な発話が行われている状態に相当する場合には、図12(C)に示すように、注目フレームから過去方向および未来方向のいずれにも、定常的な音声に相当する合成音データが存在すると考えられる。そして、このような場合には、ラグ対応の過去データと、ラグ対応の未来データのうちのいずれか一方だけでなく、その両方を、予測タップやクラスタップに含めることによって、さらなる音質の向上を図ることができると考えられる。

【0142】そこで、図11のタップ生成部301と302は、合成音データの波形の推移が、例えば、図12(A)乃至図12(C)に示したうちのいずれであるかを判定し、その判定結果に基づいて、予測タップとクラスタップを、それぞれ生成するようになっている。

【0143】即ち、図13は、図11のタップ生成部301の構成例を示している。

【0144】合成音メモリ311には、音声合成フィルタ29(図11)が出力する合成音データが順次供給されるようになっており、合成音メモリ311は、その合成音データを、順次記憶する。なお、合成音メモリ311は、注目データとされる合成音データについて予測タップとされる可能性のある合成音データのうちの、最も過去のサンプルから、最も未来のサンプルまでの間の合成音データを記憶することのできる記憶容量を少なくとも有している。また、合成音メモリ311は、その記憶容量分だけ、合成音データを記憶すると、次に供給される合成音データを、最も古い記憶値に上書きする形で記憶するようになっている。

【0145】Lコードメモリ312には、チャンネルデコード21(図11)が出力するサブフレーム単位のLコードが順次供給されるようになっており、Lコードメモリ312は、そのLコードを、順次記憶する。なお、Lコードメモリ312は、注目データとされる合成音データについて予測タップとされる可能性のある合成音データのうちの、最も過去のサンプルが配置されるサブフレームから、最も未来のサンプルが配置されるサブフレームまでの間のLコードを記憶することのできる記憶容量を少なくとも有しており、また、その記憶容量分だけ、Lコードを記憶すると、次に供給されるLコードを、最も古い記憶値に上書きする形で記憶するようになっている。

【0146】フレームパワー計算部313は、合成音メモリ311に記憶された合成音データを用い、所定のフレーム単位で、そのフレームにおける合成音データのパワーを求め、バッファ314に供給する。なお、フレームパワー計算部313でパワーを求める単位であるフレームは、CELP方式におけるフレームやサブフレームに一致していても良いし、一致していなくても良い。従って、フレームパワー計算部313でパワーを求める単位であるフレームは、CELP方式におけるフレームを構成する160サンプルや、サブフレームを構成する40サンプル以外の値である、例えば、128サンプル等で構成することが可能である。但し、本実施の形態では、説明を簡単にするため、フレームパワー計算部313でパワーを求める単位であるフレームは、CELP方式におけるフレームと一致するものとする。

【0147】バッファ314は、フレームパワー計算部313から順次供給される合成音データのパワーを、順次記憶する。なお、バッファ314は、少なくとも、注

目フレーム、およびその直前と直後のフレームの、合計で3フレーム分の合成音データのパワーを記憶することができるようになっており、また、その記憶容量分だけ、パワーを記憶すると、次にフレームパワー計算部313から供給されるパワーを、最も古い記憶値に上書きする形で記憶するようになっている。

【0148】状態判定部315は、バッファ314に記憶されたパワーに基づき、注目データ付近の合成音データの波形の推移を判定する。即ち、状態判定部315は、注目データ付近の合成音データの波形の推移が、図12(A)に示したように、注目フレームの直前のフレームが無音の状態(以下、適宜、立ち上がり状態という)、図12(B)に示したように、注目フレームの直後のフレームが無音の状態(以下、適宜、たち下がり状態という)、または図12(C)に示したように、注目フレームの直前から直後にかけて定常になっている状態(以下、適宜、定常状態という)のうちのいずれの状態になっているかを判定する。そして、状態判定部315は、その判定結果を、データ抽出部316に供給する。

【0149】データ抽出部316は、合成音メモリ311から、注目サブフレームの合成音データを読み出すことにより抽出する。さらに、データ抽出部316は、状態判定部315からの波形の推移の判定結果に基づき、Lコードメモリ312を参照して、合成音メモリ311から、ラグ対応の過去データもしくはラグ対応の未来データのうちの一方または両方を読み出すことにより抽出する。そして、データ抽出部316は、合成音メモリ311から読み出した注目サブフレームの合成音データと、ラグ対応の過去データもしくはラグ対応の未来データのうちの一方または両方とを、予測タップとして出力する。

【0150】次に、図14のフローチャートを参照して、図13のタップ生成部301の処理について説明する。

【0151】合成音メモリ311には、音声合成フィルタ29(図11)が出力する合成音データが順次供給され、合成音メモリ311は、その合成音データを、順次記憶する。また、Lコードメモリ312には、チャンネルデコード21(図11)が出力するサブフレーム単位のLコードが順次供給され、Lコードメモリ312は、そのLコードを、順次記憶する。

【0152】一方、フレームパワー計算部313は、合成音メモリ311に記憶された合成音データを、フレーム単位で順次読み出し、各フレームにおける合成音データのパワーを求めて、バッファ314に記憶させている。

【0153】そして、状態判定部315は、ステップS21において、バッファ314から、注目フレームのパワー P_n 、その直前のフレームのパワー P_{n-1} 、およびその直後のフレームのパワー P_{n+1} を読み出し、注目フレ

ームのパワー P_{n-1} と、直前のフレームのパワー P_n との差分値 $P_n - P_{n-1}$ を計算するとともに、直後のフレームのパワー P_{n+1} と、注目フレームのパワー P_n との差分値 $P_{n+1} - P_n$ とを計算して、ステップ S 22 に進む。

【0154】ステップ S 22 では、状態判定部 315 は、差分値 $P_n - P_{n-1}$ の絶対値、および差分値 $P_{n+1} - P_n$ の絶対値が、いずれも、所定の閾値 ϵ よりも大であるか（以上であるか）どうかを判定する。

【0155】ステップ S 22 において、差分値 $P_n - P_{n-1}$ の絶対値、または差分値 $P_{n+1} - P_n$ の絶対値のうち 10 の少なくとも一方が、所定の閾値 ϵ より大でないと判定された場合、状態判定部 315 は、注目データ付近の合成音データの波形の推移が、図 12 (C) に示したように、注目フレームの直前から直後にかけて定常になっている定常状態になっていると判定し、その旨を表す「定常状態」メッセージを、データ抽出部 316 に供給して、ステップ S 23 に進む。

【0156】ステップ S 23 では、データ抽出部 316 は、状態判定部 315 から「定常状態」メッセージを受信すると、合成音メモリ 311 から、注目サブフレーム 20 の合成音データを読み出すとともに、さらに、L コードメモリ 312 を参照して、タグ対応の過去データおよびタグ対応の未来データとしての合成音データを読み出す。そして、データ抽出部 316 は、これらの合成音データを予測タップとして出力し、処理を終了する。

【0157】また、ステップ S 22 において、差分値 $P_n - P_{n-1}$ の絶対値、および差分値 $P_{n+1} - P_n$ の絶対値が、いずれも、所定の閾値 ϵ より大であると判定された場合、ステップ S 24 に進み、状態判定部 315 は、差分値 $P_n - P_{n-1}$ と、差分値 $P_{n+1} - P_n$ が、いずれも正であるかどうかを判定する。ステップ S 24 において、差分値 $P_n - P_{n-1}$ と、差分値 $P_{n+1} - P_n$ が、いずれも正であると判定された場合、状態判定部 315 は、注目データ付近の合成音データの波形の推移が、図 12 (A) に示したように、注目フレームの直前のフレームが無音の状態である立ち上がり状態になっていると判定し、その旨を表す「立ち上がり状態」メッセージを、データ抽出部 316 に供給して、ステップ S 25 に進む。

【0158】ステップ S 25 では、データ抽出部 316 は、状態判定部 315 から「立ち上がり状態」メッセージを受信すると、合成音メモリ 311 から、注目サブフレームの合成音データを読み出すとともに、さらに、L コードメモリ 312 を参照して、タグ対応の未来データとしての合成音データを読み出す。そして、データ抽出部 316 は、これらの合成音データを予測タップとして出力し、処理を終了する。

【0159】一方、ステップ S 24 において、差分値 $P_n - P_{n-1}$ と、差分値 $P_{n+1} - P_n$ のうちの少なくとも一方が、正でないと判定された場合、ステップ S 26 に進み、状態判定部 315 は、差分値 $P_n - P_{n-1}$ と、差分値

$P_{n+1} - P_n$ が、いずれも負であるかどうかを判定する。ステップ S 26 において、差分値 $P_n - P_{n-1}$ と、差分値 $P_{n+1} - P_n$ のうちの少なくとも一方が、負でないと判定された場合、状態判定部 315 は、注目データ付近の合成音データの波形の推移が、定常状態になっていると判定し、その旨を表す「定常状態」メッセージを、データ抽出部 316 に供給して、ステップ S 23 に進む。

【0160】ステップ S 23 では、データ抽出部 316 は、上述したように、合成音メモリ 311 から、注目サブフレームの合成音データと、タグ対応の過去データおよびタグ対応の未来データを読み出し、予測タップとして出力して、処理を終了する。

【0161】また、ステップ S 26 において、差分値 $P_n - P_{n-1}$ と、差分値 $P_{n+1} - P_n$ が、いずれも負であると判定された場合、状態判定部 315 は、注目データ付近の合成音データの波形の推移が、図 12 (B) に示したように、注目フレームの直後のフレームが無音の状態である立ち下がり状態になっていると判定し、その旨を表す「立ち下がり状態」メッセージを、データ抽出部 316 に供給して、ステップ S 27 に進む。

【0162】ステップ S 27 では、データ抽出部 316 は、状態判定部 315 から「立ち下がり状態」メッセージを受信すると、合成音メモリ 311 から、注目サブフレームの合成音データを読み出すとともに、さらに、L コードメモリ 312 を参照して、タグ対応の過去データとしての合成音データを読み出す。そして、データ抽出部 316 は、これらの合成音データを予測タップとして出力し、処理を終了する。

【0163】なお、図 11 のタップ生成部 302 も、図 13 に示したタップ生成部 301 と同様に構成することができ、この場合、図 14 で説明したようにして、クラスタップを構成させることができる。但し、図 13 において、合成音メモリ 311、L コードメモリ 312、フレームパワー計算部 313、バッファ 314、および状態判定部 315 は、タップ生成部 301 と 302 とで兼用することが可能である。

【0164】また、上述の場合には、注目フレームと、その直前または直後のフレームそれぞれにおけるパワーを比較して、注目データ付近の合成音データの波形の推移を判定するようにしたが、注目データ付近の合成音データの波形の推移を判定は、その他、例えば、注目フレームと、より過去または未来に亘るフレームそれぞれにおけるパワーを比較して行うことも可能である。

【0165】さらに、上述の場合には、注目データ付近の合成音データの波形の推移を、「定常状態」、「立ち上がり状態」、または「立ち下がり状態」の 3 つの状態のうちのいずれかに判定するようにしたが、4 以上の状態のいずれかに判定することも可能である。即ち、例えば、図 14 では、ステップ S 22 において、差分値 $P_n - P_{n-1}$ の絶対値、および差分値 $P_{n+1} - P_n$ の絶対値

を、いずれも、1つの閾値 ε と比較し、その大小関係を判定しているが、差分値 $P_{n+1}-P_n$ の絶対値や、差分値 $P_{n+1}-P_n$ の絶対値を、複数の閾値と比較するようにすることで、注目データ付近の合成音データの波形の推移を、4以上の状態のいずれかに判定することが可能である。

【0166】そして、このように、注目データ付近の合成音データの波形の推移を、4以上の状態のいずれかに判定する場合においては、予測タップは、注目サブフレームの合成音データと、タグ対応の過去データまたはタグ対応の未来データの他、例えば、そのタグ対応の過去データまたはタグ対応の未来データを注目データとしたときに、タグ対応の過去データまたはタグ対応の未来データとなる合成音データを含めて構成すること等が可能である。

【0167】ところで、タップ生成部301において、以上のように予測タップを生成する場合には、その予測タップを構成する合成音データのサンプル数は変化する。このことは、タップ生成部302において生成されるクラスタップについても同様である。

【0168】予測タップについては、それを構成するデータの数(タップ数)が変化しても、その予測タップと同一の数のタップ係数を、後述する図16の学習装置において学習し、係数メモリ124に記憶させておけば良いから問題はない。

【0169】一方、クラスタップについては、それを構成するタップ数が増加すると、各タップ数のクラスタップごとに得られる全クラス数が増加することから、処理が複雑になるおそれがある。そこで、クラスタップのタップ数が増加しても、そのクラスタップにより得られるクラス数が増加しないようなクラス分類を行うのが望ましい。

【0170】このように、クラスタップのタップ数が増加しても、そのクラスタップにより得られるクラス数が増加しないようなクラス分類を行う方法としては、クラス分類において、例えば、クラスタップの構成を考慮する方法がある。

【0171】即ち、本実施の形態では、クラスタップが、注目サブフレームの合成音データの他に、タグ対応の過去データもしくはタグ対応の未来データ的一方または両方を含めて構成されることによって、クラスタップのタップ数が増減する。そこで、例えば、いま、クラスタップが、注目サブフレームの合成音データと、タグ対応の過去データもしくはタグ対応の未来データ的一方から構成される場合に、そのタップ数がS個であるとし、クラスタップが、注目サブフレームの合成音データと、タグ対応の過去データおよびタグ対応の未来データの両方から構成される場合に、そのタップ数がL(>S)個であるとする。そして、タップ数がS個の場合

の場合は、 $n+m$ ビットのクラスコードが得られるとする。

【0172】この場合、クラスコードとして、 $n+m+2$ ビットを用いるとともに、その $n+m+2$ ビットのうちの、例えば、上位ビットの2ビットを、クラスタップが、タグ対応の過去データを含む場合、タグ対応の未来データを含む場合、その両方を含む場合の3つの場合とで、それぞれ、例えば、“00”、“01”、“10”に設定することにより、タップ数がS個とL個のいずれであっても、全クラス数が 2^{n+m+2} クラスのクラス分類が可能となる。

【0173】即ち、クラスタップが、タグ対応の過去データと、タグ対応の未来データの両方を含み、そのタップ数がL個の場合には、 $n+m$ ビットのクラスコードが得られるクラス分類を行い、その $n+m$ ビットのクラスコードに、その上位2ビットとして、クラスタップが、タグ対応の過去データと、タグ対応の未来データの両方を含むことを表す“10”を付加した $n+m+2$ ビットを最終的なクラスコードとすれば良い。

【0174】また、クラスタップが、タグ対応の過去データを含み、そのタップ数がS個の場合には、 n ビットのクラスコードが得られるクラス分類を行い、その n ビットのクラスコードに、その上位ビットとして、 m ビットの“0”を付加して $n+m$ ビットとし、さらに、その $n+m$ ビットに、上位ビットとして、クラスタップが、タグ対応の過去データを含むことを表す“00”を付加した $n+m+2$ ビットを最終的なクラスコードとすれば良い。

【0175】さらに、クラスタップが、タグ対応の未来データを含み、そのタップ数がS個の場合には、 n ビットのクラスコードが得られるクラス分類を行い、その n ビットのクラスコードに、その上位ビットとして、 m ビットの“0”を付加して $n+m$ ビットとし、さらに、その $n+m$ ビットに、上位ビットとして、クラスタップが、タグ対応の未来データを含むことを表す“01”を付加した $n+m+2$ ビットを最終的なクラスコードとすれば良い。

【0176】次に、図13のタップ生成部301では、フレームパワー計算部313において、合成音データから、そのフレーム単位のパワーを計算するようにした

が、CELP方式によって、音声符号化した符号化データ(コードデータ)には、上述したように、フレームエネルギーが含まれる場合があり、この場合には、そのフレームエネルギーを、そのフレームにおける合成音のパワーとして採用することが可能である。

【0177】図15は、フレームエネルギーを、そのフレームにおける合成音のパワーとして採用する場合の、図11のタップ生成部301の構成例を示している。なお、図中、図13における場合と対応する部分については、同一の符号を付してある。即ち、図15のタップ生

10

20

30

40

50

成部 301 は、フレームパワー計算部 313 が設けられていない他は、図 13 における場合と同様に構成されている。

【0178】但し、バッファ 314 には、受信部 114 (図 11) に供給される符号化データ (コードデータ) に含まれる、フレームごとのフレームエネルギーが供給されているようになっており、バッファ 314 は、このフレームエネルギーを記憶するようになっている。そして、状態判定部 315 は、このフレームエネルギーを、上述した、合成音データから求められるフレーム単位のパワーと同様に用いて、注目データ付近の合成音データの波形の推移を判定する。

【0179】ここで、符号化データに含まれる、フレームごとのフレームエネルギーは、チャンネルエンコーダ 21 において、符号化データから分離され、タップ生成部 301 に供給される。

【0180】なお、タップ生成部 302 も、図 15 に示したように構成することが可能である。

【0181】次に、図 16 は、受信部 114 が図 11 に示したように構成される場合に、その係数メモリ 24 に記憶されるタップ係数を学習する学習装置の一実施の形態の構成例を示している。なお、図中、図 9 における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。即ち、図 16 の学習装置は、タップ生成部 131 と 132 に替えて、タップ生成部 321 と 322 がそれぞれ設けられている他は、図 9 における場合と同様に構成される。

【0182】タップ生成部 321 と 322 は、図 11 のタップ生成部 301 と 302 における場合とそれぞれ同様にして、予測タップとクラスタップを構成する。

【0183】従って、この場合、より高音質の音声を復号することのできるタップ係数を得ることができる。

【0184】なお、学習装置において、予測タップやクラスタップの生成にあたり、注目データ付近の合成音データの波形の推移の判定を、図 15 で説明したように、フレームごとのフレームエネルギーを用いて行う場合には、そのフレームエネルギーは、LPC 分析部 204 における LPC 分析の過程で得られる自己相関係数を用いて計算することができる。

【0185】そこで、図 17 は、自己相関係数からフレームエネルギーを求める場合の図 16 のタップ生成部 321 の構成例を示している。なお、図中、図 13 のタップ生成部 301 における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。即ち、図 17 のタップ生成部 321 は、フレームパワー計算部 313 に替えてフレームエネルギー計算部 331 が設けられている他は、図 13 におけるタップ生成部 301 と同様に構成されている。

【0186】フレームエネルギー計算部 331 には、図 16 の LPC 分析部 204 が LPC 分析を行う過程で求め

られる音声の自己相関係数が供給されるようになっており、フレームエネルギー計算部 331 は、その自己相関係数から、符号化データ (コードデータ) に含まれるフレームエネルギーを計算し、バッファ 314 に供給する。

【0187】従って、図 17 の実施の形態では、状態判定部 315 は、このフレームエネルギーを、上述した、合成音データから求められるフレーム単位のパワーと同様に用いて、注目データ付近の合成音データの波形の推移を判定する。

【0188】なお、図 16 のクラスタップを生成するタップ生成部 322 も、図 17 に示したように構成することが可能である。

【0189】次に、図 18 は、図 4 の受信部 114 の第 3 の構成例を示している。なお、図中、図 5 または図 11 における場合と対応する部分については、同一の符号を付してあり、その説明は、適宜省略する。

【0190】図 5 や図 11 の受信部 114 は、音声合成フィルタ 29 が出力する合成音データに対して、クラス分類適応処理を施すことによって、高音質の音声を復号するようになっていたが、図 18 の受信部 114 は、音声合成フィルタ 29 に入力される残差信号 (復号残差信号) と、線形予測係数 (復号線形予測係数) に対して、クラス分類適応処理を施すことによって、高音質の音声を復号するようになっている。

【0191】即ち、適応コードブック記憶部 22、ゲイン復号器 23、励起コードブック記憶部 24、および演算器 26 乃至 28 において、Lコード、Gコード、および Iコードから復号される残差信号である復号残差信号や、フィルタ係数復号器 25 において、Aコードから復号される線形予測係数である復号線形予測係数は、前述したように、誤差を含むものとなっており、そのまま音声合成フィルタ 29 に入力したのでは、音声合成フィルタ 29 から出力される合成音データの音質が劣化する。

【0192】そこで、図 18 の受信部 114 では、学習により求めたタップ係数を用いた予測演算を行うことにより、真の残差信号と線形予測係数の予測値を求め、これらを音声合成フィルタ 29 に与えることで、高音質の合成音を生成するようになっている。

【0193】即ち、図 18 の受信部 114 では、例えば、クラス分類適応処理を利用して、復号残差信号が、真の残差信号 (の予測値) に復号されるとともに、復号線形予測係数が、真の線形予測係数 (の予測値) に復号され、その残差信号と線形予測係数を、音声合成フィルタ 29 に与えることで、高音質の合成音データが求められる。

【0194】そのために、演算器 28 が出力する復号残差信号は、タップ生成部 341 と 32 に供給される。また、タップ生成部 341 と 342 には、チャンネルデコーダ 21 が出力する Lコードも供給されるようになっている。

10

20

30

40

50

【0195】そして、タップ生成部341は、図5のタップ生成部121や図11のタップ生成部301と同様に、そこに供給される復号残差信号から、予測タップとするサンプルを、Lコードに基づいて抽出し、予測部345に供給する。

【0196】タップ生成部342も、図5のタップ生成部122や図11のタップ生成部302と同様に、そこに供給される復号残差信号から、クラスタップとするサンプルを、Lコードに基づいて抽出し、予測部345に供給する。

【0197】クラス分類部343は、タップ生成部342から供給されるクラスタップに基づいて、クラス分類を行い、そのクラス分類結果としてのクラスコードを、係数メモリ344に供給する。

【0198】係数メモリ344は、後述する図21の学習装置において学習処理が行われることにより得られる、クラスごとの残差信号についてのタップ係数 $w_{(c)}$ を記憶しており、クラス分類部343が出力するクラスコードに対応するアドレスに記憶されているタップ係数を、予測部345に供給する。

【0199】予測部345は、タップ生成部341が出力する予測タップと、係数メモリ344が出力する残差信号についてのタップ係数とを取得し、その予測タップとタップ係数とを用いて、式(6)に示した線形予測演算を行う。これにより、予測部345は、注目サブフレームの残差信号(の予測値) e_m を求めて、音声合成フィルタ29に、入力信号として供給する。

【0200】タップ生成部351と352には、フィルタ係数復号器25が出力する、サブフレームごとの復号線形予測係数 α_c が供給されるようになっており、タップ生成部351と352は、その復号線形予測係数から、予想タップとクラスタップとするものをそれぞれ抽出する。ここで、タップ生成部351と352は、例えば、注目サブフレームの線形予測係数すべてを、それぞれ、予測タップとクラスタップとする。予測タップは、タップ生成部351から予測部355に供給され、クラスタップは、タップ生成部352からクラス分類部353に供給される。

【0201】クラス分類部353は、タップ生成部352から供給されるクラスタップに基づいて、クラス分類を行い、そのクラス分類結果としてのクラスコードを、係数メモリ354に供給する。

【0202】係数メモリ354は、後述する図21の学習装置において学習処理が行われることにより得られる、クラスごとの線形予測係数についてのタップ係数 $w_{(c)}$ を記憶しており、クラス分類部353が出力するクラスコードに対応するアドレスに記憶されているタップ係数を、予測部355に供給する。

【0203】予測部355は、タップ生成部351が出力する予測タップと、係数メモリ354が出力する線形

予測係数についてのタップ係数とを取得し、その予測タップとタップ係数とを用いて、式(6)に示した線形予測演算を行う。これにより、予測部355は、注目サブフレームの線形予測係数(の予測値) m_{α_c} を求めて、音声合成フィルタ29に供給する。

【0204】次に、図19のフローチャートを参照して、図18の受信部114の処理について説明する。

【0205】チャンネルデコーダ21は、そこに供給されるコードデータから、Lコード、Gコード、Iコード、Aコードを分離し、それぞれを、適応コードブック記憶部22、ゲイン復号器23、励起コードブック記憶部24、フィルタ係数復号器25に供給する。さらに、Lコードは、タップ生成部341と342にも供給される。

【0206】そして、適応コードブック記憶部22、ゲイン復号器23、励起コードブック記憶部24、演算器26乃至28では、図1の適応コードブック記憶部9、ゲイン復号器10、励起コードブック記憶部11、演算器12乃至14における場合と同様の処理が行われ、これにより、Lコード、Gコード、およびIコードが、残差信号 e に復号される。この復号残差信号は、演算器28からタップ生成部341と342に供給される。

【0207】さらに、フィルタ係数復号器25は、図2で説明したように、そこに供給されるAコードを、復号線形予測係数に復号し、タップ生成部351と352に供給する。

【0208】そして、ステップS31において、予測タップとクラスタップが生成される。

【0209】即ち、タップ生成部341は、そこに供給される復号残差信号のサブフレームを、順次、注目サブフレームとし、さらに、その注目サブフレームの復号残差信号のサンプル値を、順次、注目データとして、注目サブフレームにおける復号残差信号を抽出するとともに、チャンネルデコーダ21が出力する注目サブフレームに配置されたLコード等に基づいて、注目サブフレーム以外の復号残差信号を抽出し、即ち、注目サブフレームに配置されているLコードが表すラグだけ過去の位置を始点とする40サンプルの復号残差信号(これも、以下、適宜、ラグ対応の過去データという)、またはLコードが表すラグだけ過去の位置が注目データの位置となるLコードが配置された、注目サブフレームから見て未来方向のサブフレームに配置された40サンプルの復号残差信号(これも、以下、適宜、ラグ対応の未来データという)を抽出し、予測タップを生成する。タップ生成部342も、タップ生成部341と同様に、クラスタップを生成する。

【0210】さらに、ステップS31では、タップ生成部351と352が、フィルタ係数復号器25が出力する注目サブフレームの復号線形予測係数を、それぞれ、予測タップとクラスタップとして抽出する。

【0211】そして、タップ生成部341で得られた予

測タップは予測部 345 に、タップ生成部 342 で得られたクラスタップはクラス分類部 343 に、タップ生成部 351 で得られた予測タップは予測部 355 に、タップ生成部 352 で得られたクラスタップはクラス分類部 353 に、それぞれ供給される。

【0212】そして、ステップ S32 に進み、クラス分類部 343 が、タップ生成部 342 から供給されるクラスタップに基づいて、クラス分類を行い、その結果得られるクラスコードを、係数メモリ 344 に供給するとともに、クラス分類部 353 が、タップ生成部 352 から供給されるクラスタップに基づいて、クラス分類を行い、その結果得られるクラスコードを、係数メモリ 354 に供給して、ステップ S33 に進む。

【0213】ステップ S33 では、係数メモリ 344 が、クラス分類部 343 から供給されるクラスコードに対応するアドレスから、残差信号についてのタップ係数を読み出し、予測部 345 に供給するとともに、係数メモリ 354 が、クラス分類部 343 から供給されるクラスコードに対応するアドレスから、線形予測係数についてのタップ係数を読み出し、予測部 355 に供給する。

【0214】そして、ステップ S34 に進み、予測部 345 は、係数メモリ 344 が出力する残差信号についてのタップ係数を取得し、そのタップ係数と、タップ生成部 341 からの予測タップとを用いて、式 (6) に示した積和演算を行い、注目サブフレームの真の残差信号

(の予測値)を得る。さらに、ステップ S34 では、予測部 355 は、係数メモリ 344 が出力する線形予測係数についてのタップ係数を取得し、そのタップ係数と、タップ生成部 351 からの予測タップとを用いて、式 (6) に示した積和演算を行い、注目サブフレームの真の線形予測係数 (の予測値)を得る。

【0215】以上のようにして得られた残差信号および線形予測係数は、音声合成フィルタ 29 に供給され、音声合成フィルタ 29 では、その残差信号および線形予測係数を用いて、式 (4) の演算が行われることにより、注目サブフレームの注目データに対応する合成音データが生成される。この合成音データは、音声合成フィルタ 29 から、D/A 変換部 30 を介して、スピーカ 31 に供給され、これにより、スピーカ 31 からは、その合成音データに対応する合成音出力される。

【0216】予測部 345 と 355 において、残差信号と線形予測係数がそれぞれ得られた後は、ステップ S35 に進み、まだ、注目サブフレームとして処理すべきサブフレームの L コード、G コード、I コード、および A コードがあるかどうか判定される。ステップ S35 において、まだ、注目サブフレームとして処理すべきフレームの L コード、G コード、I コード、および A コードがあると判定された場合、ステップ S31 に戻り、次に注目サブフレームとすべきサブフレームを、新たに注目サブフレームとして、以下、同様の処理を繰り返す。ま

た、ステップ S35 において、注目サブフレームとして処理すべきフレームの L コード、G コード、I コード、および A コードがないと判定された場合、処理を終了する。

【0217】次に、図 18 のタップ生成部 341 においては (クラスタップを生成するタップ生成部 342 についても同様)、予測タップが、注目サブフレームの復号残差信号と、ラグ対応の過去データもしくはラグ対応の未来データのうちの一方または両方から構成されるが、その構成は、固定とすることもできるが、残差信号の波形の推移に基づいて可変にすることもできる。

【0218】図 20 は、残差信号の波形の推移に基づいて、予測タップの構成を可変にする場合のタップ生成部 341 の構成例を示している。なお、図中、図 13 における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。即ち、図 20 のタップ生成部 341 は、合成音メモリ 311 とフレームパワー計算部 313 に替えて、残差信号メモリ 361 とフレームパワー計算部 363 が設けられている。他は、図 13 のタップ生成部 301 と同様に構成されている。

【0219】残差信号メモリ 361 には、演算器 28 (図 18) が出力する復号残差信号が順次供給されるようになっており、残差信号メモリ 361 は、その復号残差信号を、順次記憶する。なお、残差信号メモリ 361 は、注目データについて予測タップとされる可能性のある復号残差信号のうちの、最も過去のサンプルから、最も未来のサンプルまでの間の復号残差信号を記憶することのできる記憶容量を少なくとも有している。また、残差信号メモリ 361 は、その記憶容量分だけ、復号残差信号を記憶すると、次に供給される復号残差信号のサンプル値を、最も古い記憶値に上書きする形で記憶するようになっている。

【0220】フレームパワー計算部 363 は、残差信号メモリ 361 に記憶された残差信号を用い、所定のフレーム単位で、そのフレームにおける残差信号のパワーを求め、バッファ 314 に供給する。なお、フレームパワー計算部 363 でパワーを求める単位であるフレームは、図 13 のフレームパワー計算部 313 における場合と同様に、CELP 方式におけるフレームやサブフレームに一致していても良いし、一致していなくても良い。

【0221】従って、図 20 のタップ生成部 341 では、合成音データのパワーでなく、復号残差信号のパワーが求められ、そのパワーに基づいて、残差信号の波形の推移が、例えば、図 12 で説明したように、立ち上がり状態、たち下がり状態、定常状態のうちのいずれであるかが判定される。そして、その判定結果に基づいて、注目サブフレームの復号残差信号の他、ラグ対応の過去データもしくはラグ対応の未来データのうちの一方または両方が抽出され、予測タップが生成される。

【0222】なお、図18のタップ生成部342も、図20に示したタップ生成部341と同様に構成することができる。

【0223】また、図18の実施の形態では、復号残差信号についてだけ、Lコードに基づいて、予測タップやクラスタップを生成するようにしたが、復号線形予測係数についても、Lコードに基づいて、注目サブフレーム以外の復号線形予測係数を抽出し、予測タップやクラスタップを生成するようにすることが可能である。この場合、図18において点線で示すように、チャンネルデコード21が出力するLコードを、タップ生成部351や352に供給するようにすれば良い。

【0224】さらに、上述の場合においては、予測タップやクラスタップを、合成音データから生成するときには、合成音データのパワーを求め、そのパワーに基づき、合成音データの波形の推移を判定し、また、復号残差信号から生成するときには、復号残差信号のパワーを求め、そのパワーに基づき、合成音データの波形の推移を判定するようにしたが、合成音データの波形の推移は、残差信号のパワーに基づいて判定することが可能であり、同様に、残差信号の波形の推移は、合成音データのパワーに基づいて判定することが可能である。

【0225】次に、図21は、図18の係数メモリ344と354に記憶させるタップ係数の学習処理を行う学習装置の一実施の形態の構成例を示している。なお、図中、図16における場合と対応する部分については、同一の符号を付してあり、以下では、その説明は、適宜省略する。

【0226】予測フィルタ370には、A/D変換部202が出力する、ディジタル信号とされた学習用の音声信号と、LPC分析部204が出力する線形予測係数が供給される。また、タップ生成部371と372には、演算器214が出力する復号残差信号（音声合成フィルタ206に供給されるのと同じ残差信号）、およびコード決定部215が出力するLコードが供給され、タップ生成部381と382には、ベクトル量子化部205が出力する復号線形予測係数（ベクトル量子化に用いられるコードブックのコードベクトル（セントロイドベクトル）を構成する線形予測係数）が供給される。さらに、正規方程式加算回路384には、LPC分析部204が出力する線形予測係数が供給される。

【0227】予測フィルタ370は、A/D変換部202から供給される学習用の音声信号のサブフレームを、順次、注目サブフレームとして、その注目サブフレームの音声信号と、LPC分析部204から供給される線形予測係数を用いて、例えば、式(1)にしたがった演算を行うことにより、注目フレームの残差信号を求める。この残差信号は、教師データとして、正規方程式加算回路374に供給される。

【0228】タップ生成部371は、演算器214から

供給される復号残差信号を用い、コード決定部215が出力するLコードに基づいて、図18のタップ生成部341における場合と同一の予測タップを生成し、正規方程式加算回路374に供給する。タップ生成部372も、演算器214から供給される復号残差信号を用い、コード決定部215が出力するLコードに基づいて、図18のタップ生成部342における場合と同一のクラスタップを生成し、クラス分類部373に供給する。

【0229】クラス分類部373は、タップ生成部371から供給されるクラスタップに基づき、図18のクラス分類部343における場合と同様にのクラス分類を行い、その結果得られるクラスコードを、正規方程式加算回路374に供給する。

【0230】正規方程式加算回路374は、予測フィルタ370からの注目サブフレームの残差信号を、教師データとして受信するとともに、タップ生成部371からの予測タップを、生徒データとして受信し、その教師データおよび生徒データを対象として、クラス分類部373からのクラスコードごとに、図9や図16の正規方程式加算回路134における場合と同様の足し込みを行うことにより、各クラスについて、残差信号に関する式(13)に示した正規方程式をたてる。

【0231】タップ係数決定回路375は、正規方程式加算回路374においてクラスごとに生成された正規方程式それぞれを解くことにより、クラスごとに、残差信号についてのタップ係数をそれぞれ求め、係数メモリ376の、各クラスに対応するアドレスにそれぞれ供給する。

【0232】係数メモリ376は、タップ係数決定回路375から供給されるクラスごとの残差信号についてのタップ係数を記憶する。

【0233】タップ生成部381は、ベクトル量子化部205から供給される、コードベクトルの要素となっている線形予測係数、即ち、復号線形予測係数を用いて、図18のタップ生成部351における場合と同一の予測タップを生成し、正規方程式加算回路384に供給する。タップ生成部382も、ベクトル量子化部205から供給される復号線形予測係数を用い、図18のタップ生成部352における場合と同一のクラスタップを生成し、クラス分類部383に供給する。

【0234】なお、図18の実施の形態において、復号線形予測係数につき、Lコードに基づいて、注目サブフレーム以外の復号線形予測係数を抽出して、予測タップやクラスタップを生成する場合には、図21のタップ生成部381や382においても、同様にして、予測タップやクラスタップを生成する必要がある。この場合、タップ生成部381や382には、図21において点線で示すように、コード決定部215が出力するLコードが供給される。

【0235】クラス分類部383は、図18のクラス分

類部 353 における場合と同様に、タップ生成部 382 からのクラスタップに基づいて、クラス分類を行い、その結果得られるクラスコードを、正規方程式加算回路 384 に供給する。

【0236】正規方程式加算回路 384 は、LPC 分析部 204 からの注目サブフレームの線形予測係数を、教師データとして受信するとともに、タップ生成部 381 からの予測タップを、生徒データとして受信し、その教師データおよび生徒データを対象として、クラス分類部 383 からのクラスコードごとに、図 9 や図 16 の正規方程式加算回路 134 における場合と同様の足し込みを行うことにより、各クラスについて、線形予測係数に関する式 (13) に示した正規方程式をたてる。

【0237】タップ係数決定回路 385 は、正規方程式加算回路 384 においてクラスごとに生成された正規方程式を解くことにより、クラスごとに、線形予測係数についてのタップ係数をそれぞれ求め、係数メモリ 386 の、各クラスに対応するアドレスに供給する。

【0238】係数メモリ 386 は、タップ係数決定回路 385 から供給されるクラスごとの線形予測係数についてのタップ係数を記憶する。

【0239】なお、学習用の音声信号として用意する音声信号によっては、正規方程式加算回路 374 や 384 において、タップ係数を求めるのに必要な数の正規方程式が得られないクラスが生じる場合があり得るが、タップ係数決定回路 375 と 385 は、そのようなクラスについては、例えば、デフォルトのタップ係数を出力する。

【0240】次に、図 22 のフローチャートを参照して、図 21 の学習装置が行う、残差信号と線形予測係数それぞれについてのタップ係数を求める学習処理について説明する。

【0241】学習装置には、学習用の音声信号が供給され、ステップ S41 において、その学習用の音声信号から、教師データと生徒データが生成される。

【0242】即ち、学習用の音声信号は、マイク 201 に入力され、マイク 201 乃至コード決定部 215 は、図 1 のマイク 1 乃至コード決定部 15 における場合とそれぞれ同様の処理を行う。

【0243】その結果、LPC 分析部 204 で得られる線形予測係数は、教師データとして、正規方程式加算回路 384 に供給される。また、この線形予測係数は、予測フィルタ 370 にも供給される。さらに、演算器 214 で得られる復号残差信号は、生徒データとして、タップ生成部 371 および 372 に供給される。

【0244】また、A/D 変換部 202 が出力するデジタルの音声信号は、予測フィルタ 370 に供給され、ベクトル量子化部 205 が出力する復号線形予測係数は、生徒データとして、タップ生成部 381 および 382 に供給される。さらに、コード決定部 215 は、自乗

誤差最小判定部 208 から確定信号を受信したときの、自乗誤差最小判定部 208 からの L コードを、タップ生成部 371 および 372 に供給する。

【0245】そして、予測フィルタ 370 は、A/D 変換部 202 から供給される学習用の音声信号のサブフレームを、順次、注目サブフレームとして、その注目サブフレームの音声信号と、LPC 分析部 204 から供給される線形予測係数（注目サブフレームの音声信号から求められた線形予測係数）を用いて、式 (1) にしたがった演算を行うことにより、注目サブフレームの残差信号を求める。この予測フィルタ 370 で得られる残差信号は、教師データとして、正規方程式加算回路 374 に供給される。

【0246】以上のようにして、教師データと生徒データが得られた後は、ステップ S42 に進み、タップ生成部 371 と 372 が、演算器 214 から供給される復号残差信号を用い、コード決定部 215 からの L コードに基づいて、残差信号についての予測タップとクラスタップをそれぞれ生成する。即ち、タップ生成部 371 と 372 は、演算器 214 からの注目サブフレームの復号残差信号と、ラグ対応の過去データまたはラグ対応の未来データとから、残差信号についての予測タップとクラスタップをそれぞれ生成する。

【0247】さらに、ステップ S42 では、タップ生成部 381 と 382 が、ベクトル量子化部 205 から供給される注目サブフレームの線形予測係数から、線形予測係数についての予測タップとクラスタップを生成する。

【0248】そして、残差信号についての予測タップは、タップ生成部 371 から正規方程式加算回路 374 に供給され、残差信号についてのクラスタップは、タップ生成部 372 からクラス分類部 373 に供給される。また、線形予測係数についての予測タップは、タップ生成部 381 から正規方程式加算回路 384 に供給され、線形予測係数についてのクラスタップは、タップ生成部 382 からクラス分類回路 383 に供給される。

【0249】その後、ステップ S43 において、クラス分類部 373 と 383 が、そこに供給されるクラスタップに基づいて、クラス分類を行い、その結果得られるクラスコードを、正規方程式加算回路 384 と 374 に、それぞれ供給する。

【0250】そして、ステップ S44 に進み、正規方程式加算回路 374 が、予測フィルタ 370 からの教師データとしての注目サブフレームの残差信号、およびタップ生成部 371 からの生徒データとしての予測タップを対象として、式 (13) の行列 A とベクトル v の、上述したような足し込みを、クラス分類部 373 からのクラスコードごとに行う。さらに、ステップ S44 では、正規方程式加算回路 384 が、LPC 分析部 204 からの教師データとしての注目サブフレームの線形予測係数、およびタップ生成部 381 からの生徒データとしての予

測タップを対象として、式(13)の行列Aとベクトルvの、上述したような足し込みを、クラス分類部383からのクラスコードごとに行い、ステップS45に進む。

【0251】ステップS45では、まだ、注目サブフレームとして処理すべきフレームの学習用の音声信号があるかどうか判定される。ステップS45において、まだ、注目サブフレームとして処理すべきサブフレームの学習用の音声信号があると判定された場合、ステップS41に戻り、次のサブフレームを新たに注目サブフレームとして、以下、同様の処理が繰り返される。

【0252】また、ステップS45において、注目サブフレームとして処理すべきサブフレームの学習用の音声信号がないと判定された場合、ステップS46に進み、タップ係数決定回路375は、各クラスごとに生成された正規方程式を解くことにより、各クラスごとに、残差信号についてのタップ係数を求め、係数メモリ376の、各クラスに対応するアドレスに供給して記憶させる。さらに、タップ係数決定回路385も、各クラスごとに生成された正規方程式を解くことにより、各クラスごとに、線形予測係数についてのタップ係数を求め、係数メモリ386の、各クラスに対応するアドレスに供給して記憶させ、処理を終了する。

【0253】以上のようにして、係数メモリ376に記憶された各クラスごとの残差信号についてのタップ係数が、図18の係数メモリ344に記憶され、係数メモリ386に記憶された各クラスごとの線形予測係数についてのタップ係数が、図18の係数メモリ354に記憶されている。

【0254】従って、図18の係数メモリ344と354に記憶されたタップ係数は、それぞれ、線形予測演算を行うことにより得られる真の残差信号と線形予測係数の予測値の予測誤差(自乗誤差)が、統計的に最小になるように学習を行うことにより求められたものであるから、図18の予測部345と355が出力する残差信号と線形予測係数は、それぞれ真の残差信号と線形予測係数にほぼ一致することとなり、その結果、これらの残差信号と線形予測係数によって生成される合成音は、歪みの少ない、高音質のものとなる。

【0255】次に、上述した一連の処理は、ハードウェアにより行うこともできるし、ソフトウェアにより行うこともできる。一連の処理をソフトウェアによって行う場合には、そのソフトウェアを構成するプログラムが、汎用のコンピュータ等にインストールされる。

【0256】そこで、図23は、上述した一連の処理を実行するプログラムがインストールされるコンピュータの一実施の形態の構成例を示している。

【0257】プログラムは、コンピュータに内蔵されている記録媒体としてのハードディスク405やROM403に予め記録しておくことができる。

【0258】あるいはまた、プログラムは、フロッピー(登録商標)ディスク、CD-ROM(Compact Disc Read Only Memory)、MO(Magneto optical)ディスク、DVD(Digital Versatile Disc)、磁気ディスク、半導体メモリなどのリムーバブル記録媒体411に、一時的あるいは永続的に格納(記録)しておくことができる。このようなリムーバブル記録媒体411は、いわゆるパッケージソフトウェアとして提供することができる。

【0259】なお、プログラムは、上述したようなリムーバブル記録媒体411からコンピュータにインストールする他、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、コンピュータに無線で転送したり、LAN(Local Area Network)、インターネットといったネットワークを介して、コンピュータに有線で転送し、コンピュータでは、そのようにして転送されてくるプログラムを、通信部408で受信し、内蔵するハードディスク405にインストールすることができる。

【0260】コンピュータは、CPU(Central Processing Unit)402を内蔵している。CPU402には、バス401を介して、入出力インタフェース410が接続されており、CPU402は、入出力インタフェース410を介して、ユーザによって、キーボードや、マウス、マイク等で構成される入力部407が操作等されることにより指令が入力されると、それにしたがって、ROM(Read Only Memory)403に格納されているプログラムを実行する。あるいは、また、CPU402は、ハードディスク405に格納されているプログラム、衛星若しくはネットワークから転送され、通信部408で受信されてハードディスク405にインストールされたプログラム、またはドライブ409に装着されたリムーバブル記録媒体411から読み出されてハードディスク405にインストールされたプログラムを、RAM(Random Access Memory)404にロードして実行する。これにより、CPU402は、上述したフローチャートにしたがった処理、あるいは上述したブロック図の構成により行われる処理を行う。そして、CPU402は、その処理結果を、必要に応じて、例えば、入出力インタフェース410を介して、LCD(Liquid Crystal Display)やスピーカ等で構成される出力部406から出力、あるいは、通信部408から送信、さらには、ハードディスク405に記録等させる。

【0261】ここで、本明細書において、コンピュータに各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理(例えば、並列処理あるいはオブジェクトによる処理)も含むものである。

【0262】また、プログラムは、1のコンピュータにより処理されるものであっても良いし、複数のコンピュータによって分散処理されるものであっても良い。さら

に、プログラムは、遠方のコンピュータに転送されて実行されるものであっても良い。

【0263】なお、本実施の形態においては、学習用の音声信号として、どのようなものを用いるかについては、特に言及しなかったが、学習用の音声信号としては、人が発話した音声の他、例えば、曲（音楽）等を採用することが可能である。そして、上述したような学習処理によれば、学習用の音声信号として、人の発話を用いた場合には、そのような人の発話の音声の音質を向上させるようなタップ係数が得られ、曲を用いた場合には、曲の音質を向上させるようなタップ係数が得られることになる。

【0264】また、係数メモリ124等には、タップ係数をあらかじめ記憶させておくようにしたが、係数メモリ124等に記憶させるタップ係数は、携帯電話機101において、図3の基地局102（あるいは交換局103）や、図示しないWWW(World Wide Web)サーバ等からダウンロードするようにすることができる。即ち、上述したように、タップ係数は、人の発話用や曲用等のように、ある種類の音声信号に適したものを、学習によって得ることができる。さらに、学習に用いる教師データおよび生徒データによっては、合成音の音質に差が生じるタップ係数を得ることができる。従って、そのような各種のタップ係数を、基地局102等に記憶させておき、ユーザには、自身の所望するタップ係数をダウンロードさせるようにすることができる。そして、このようなタップ係数のダウンロードサービスは、無料で行うこともできるし、有料で行うこともできる。さらに、タップ係数のダウンロードサービスを有料で行う場合には、タップ係数のダウンロードに対する対価としての代金

は、例えば、携帯電話機101の通話料等とともに請求するようにすることが可能である。

【0265】また、係数メモリ124等は、携帯電話機101に対して着脱可能なメモリカード等で構成することができる。この場合、上述したような各種のタップ係数それぞれを記憶させた、異なるメモリカードを提供するようにすれば、ユーザは、場合に応じて、所望のタップ係数が記憶されたメモリカードを、携帯電話機101に装着して使用することが可能となる。

【0266】さらに、本発明は、例えば、VSELP(Vector Sum Excited Liner Prediction), PS-CELP(Pitch Synchronous Innovation CELP), CS-ACELP(Conjugate Structure Algebraic CELP)等のCELP方式による符号化の結果得られるコードから合成音を生成する場合に、広く適用可能である。

【0267】また、本発明は、CELP方式による符号化の結果得られるコードから合成音を生成する場合に限らず、あるコードから、残差信号と線形予測係数を得て、合成音を生成する場合に、広く適用可能である。

【0268】さらに、本発明は、音声に限らず、例え

ば、画像等にも適用可能である。即ち、本発明は、Lコードのような、周期を表す周期情報を用いて処理されるデータに、広く適用可能である。

【0269】また、本実施の形態では、タップ係数を用いた線形1次予測演算によって、高音質の音声や、残差信号、線形予測係数の予測値を求めるようにしたが、この予測値は、その他、2次以上の高次の予測演算によって求めることも可能である。

【0270】さらに、本実施の形態では、係数メモリ124等に、タップ係数自体を記憶させておくようにしたが、係数メモリ124等には、その他、例えば、無段階の調整が可能な（アナログ的な変化が可能な）タップ係数の元（種）となる情報としての係数種を記憶させておき、ユーザの操作に応じて、その係数種から、ユーザの所望する音質の音声を得られるようなタップ係数を生成するようにすることが可能である。

【0271】

【発明の効果】本発明の第1のデータ処理装置およびデータ処理方法、並びにプログラムおよび記録媒体によれば、所定のデータのうちの注目している注目データにつき、周期情報に応じて、所定のデータを抽出することにより、所定の処理に用いるタップが生成され、そのタップを用いて、注目データについて、所定の処理が行われる。従って、従って、例えば、品質の良いデータの復号等が可能となる。

【0272】本発明の第2のデータ処理装置およびデータ処理方法、並びにプログラムおよび記録媒体によれば、学習の教師となる教師データから、所定のデータと周期情報が、学習の生徒となる生徒データとして生成される。そして、生徒データとしての所定のデータのうちの注目している注目データにつき、周期情報に応じて、所定のデータを抽出することにより、教師データを予測するのに用いる予測タップが生成され、予測タップとタップ係数とを用いて、所定の予測演算を行うことにより得られる教師データの予測値の予測誤差が、統計的に最小になるように学習が行われ、タップ係数が求められる。従って、例えば、品質の良いデータを得るためのタップ係数を得ることが可能となる。

【図面の簡単な説明】

【図1】従来の携帯電話機の送信部の一例の構成を示すブロック図である。

【図2】従来の携帯電話機の受信部の一例の構成を示すブロック図である。

【図3】本発明を適用した伝送システムの一実施の形態の構成例を示す図である。

【図4】携帯電話機101、と101₂の構成例を示すブロック図である。

【図5】受信部114の第1の構成例を示すブロック図である。

【図6】図5の受信部114の処理を説明するフローチ

ャートである。

【図 7】予測タップとクラスタップの生成方法を説明する図である。

【図 8】予測タップとクラスタップの生成方法を説明する図である。

【図 9】本発明を適用した学習装置の第 1 実施の形態の構成例を示すブロック図である。

【図 10】図 9 の学習装置の処理を説明するフローチャートである。

【図 11】受信部 114 の第 2 の構成例を示すブロック図である。

【図 12】合成音データの波形の推移を示す図である。

【図 13】タップ生成部 301 と 302 の構成例を示すブロック図である。

【図 14】タップ生成部 301 と 302 の処理を説明するフローチャートである。

【図 15】タップ生成部 301 と 302 の他の構成例を示すブロック図である。

【図 16】本発明を適用した学習装置の第 2 実施の形態の構成例を示すブロック図である。

【図 17】タップ生成部 321 と 322 の構成例を示すブロック図である。

【図 18】受信部 114 の第 3 の構成例を示すブロック図である。

【図 19】図 18 の受信部 114 の処理を説明するフローチャートである。

【図 20】タップ生成部 341 と 342 の構成例を示すブロック図である。

【図 21】本発明を適用した学習装置の第 3 実施の形態の構成例を示すブロック図である。

【図 22】図 21 の学習装置の処理を説明するフローチャートである。

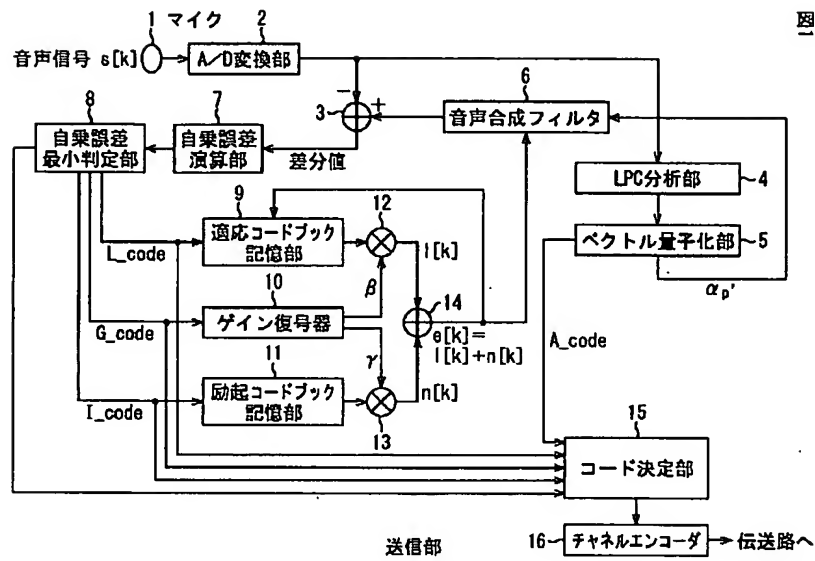
【図 23】本発明を適用したコンピュータの一実施の形態の構成例を示すブロック図である。

【符号の説明】

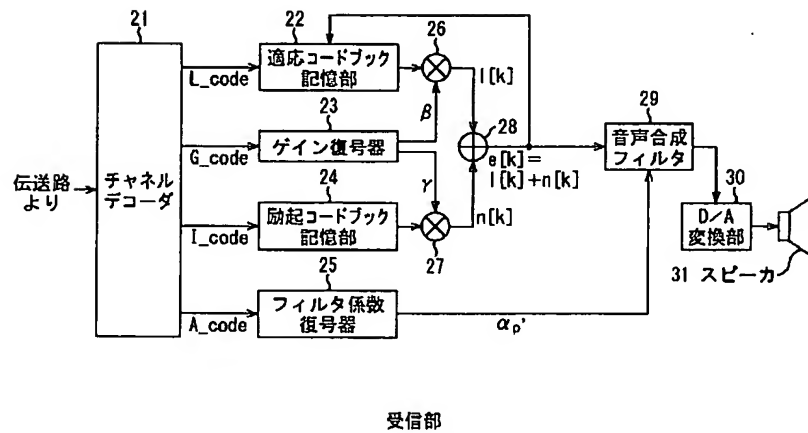
21 チャネルデコーダ、 22 適応コードブック記憶部、 23 ゲイン復号器、 24 励起コードブッ

ク記憶部、 25 フィルタ係数復号器、 26 乃至 28 演算器、 29 音声合成フィルタ、 30 D/A 変換部、 31 スピーカ、 101、 101、 携帯電話機、 102、 102、 基地局、 103 交換局、 111 アンテナ、 112 変復調部、 113 送信部、 114 受信部、 121、 122 タップ生成部、 123 クラス分類部、 124 係数メモリ、 125 予測部、 131、 132 タップ生成部、 133 クラス分類部、 134 正規方程式加算回路、 135 タップ係数決定回路、 136 係数メモリ、 201 マイク、 202 A/D 変換部、 203 演算器、 204 LPC 分析部、 205 ベクトル量子化部、 206 音声合成フィルタ、 207 自乗誤差演算部、 208 自乗誤差最小判定部、 209 適応コードブック記憶部、 210 ゲイン復号器、 211 励起コードブック記憶部、 212 乃至 214 演算器、 301、 302 タップ生成部、 311 合成音メモリ、 312 L コードメモリ、 313 フレームパワー計算部、 314 バッファ、 315 状態判定部、 316 データ抽出部、 321、 322 タップ生成部、 331 フレームエネルギー計算部、 341、 342 タップ生成部、 343 クラス分類部、 344 係数メモリ、 345 予測部、 351、 352 タップ生成部、 353 クラス分類部、 354 係数メモリ、 355 予測部、 361 残差信号メモリ、 363 フレームパワー計算部、 370 予測フィルタ、 371、 372 タップ生成部、 373 クラス分類部、 374 正規方程式加算回路、 375 タップ係数決定回路、 376 係数メモリ、 381、 382 タップ生成部、 383 クラス分類部、 384 正規方程式加算回路、 385 タップ係数決定回路、 386 係数メモリ、 401 バス、 402 CPU、 403 ROM、 404 RAM、 405 ハードディスク、 406 出力部、 407 入力部、 408 通信部、 409 ドライブ、 410 入出力インタフェース、 411 リムーバブル記録媒体

【図 1】

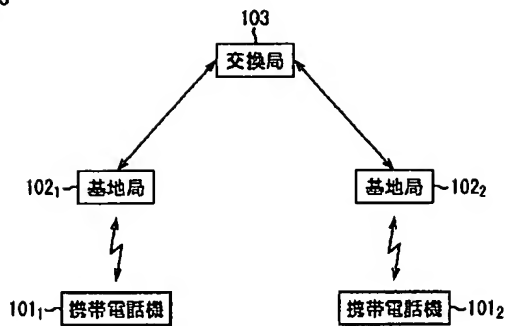


【図 2】



【図3】

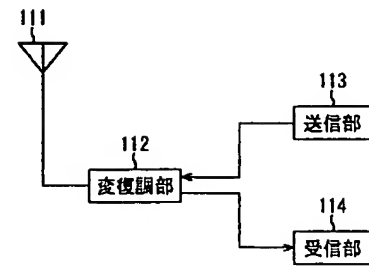
図3



伝送システム

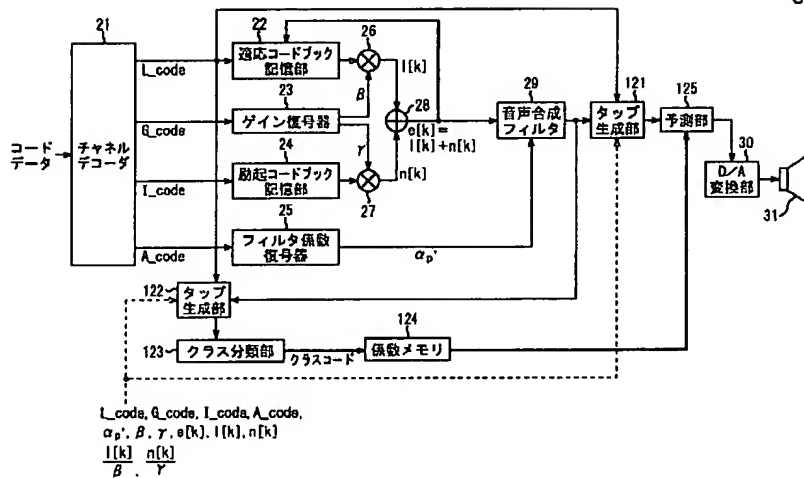
【図4】

図4



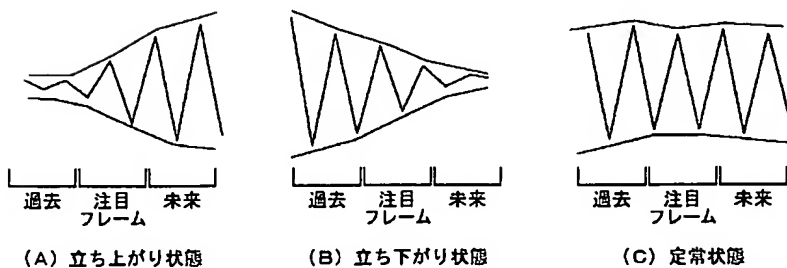
携帯電話機 101

【図5】



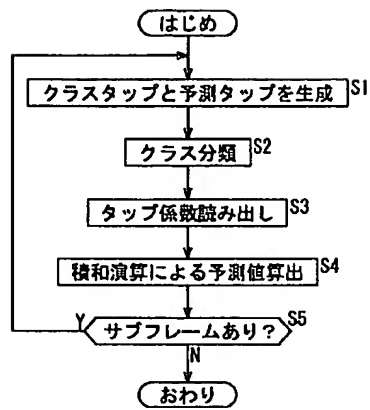
受信部 114

【図12】



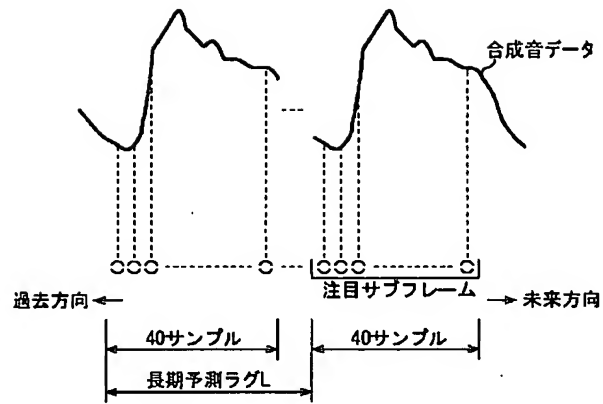
【図 6】

図 6



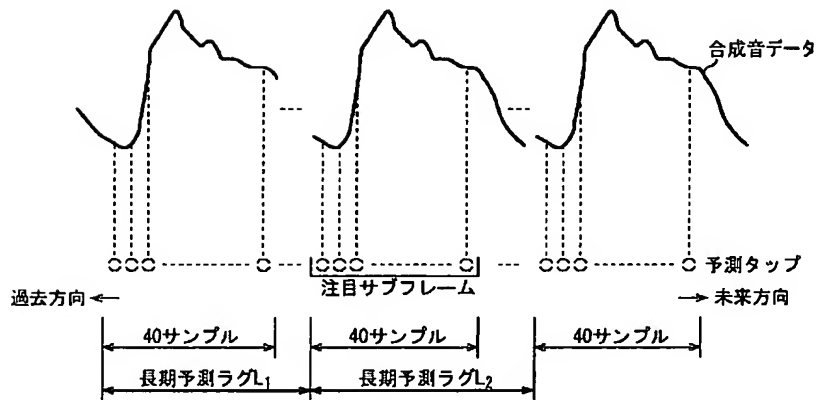
【図 7】

図 7

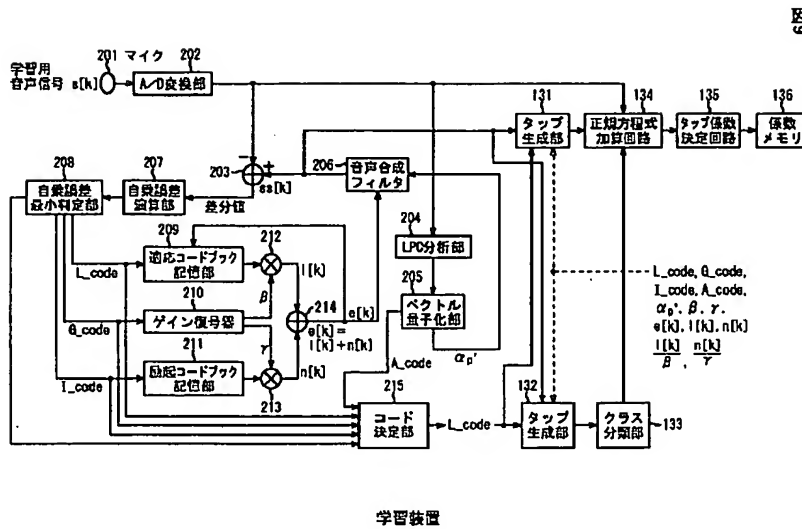


【図 8】

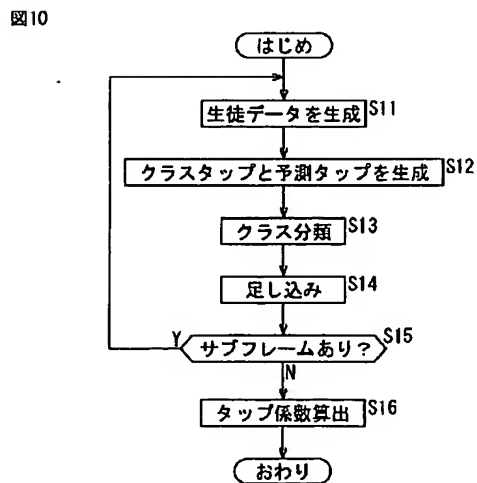
図 8



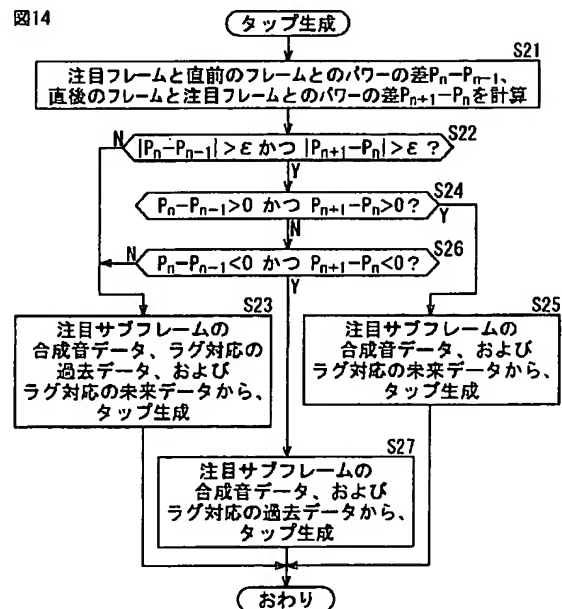
【図9】



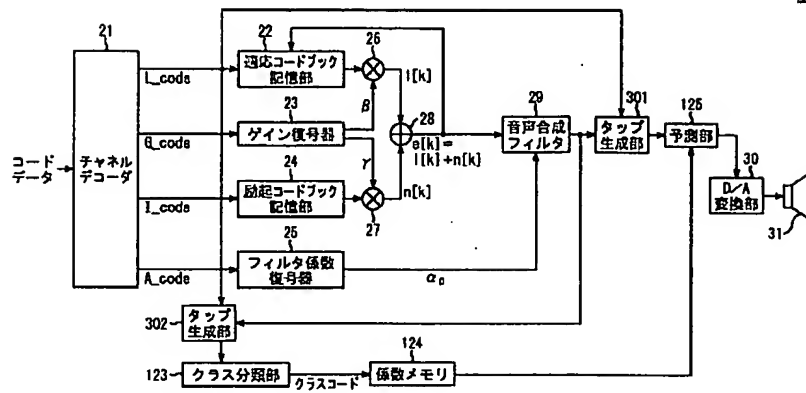
【図 10】



【図 14】

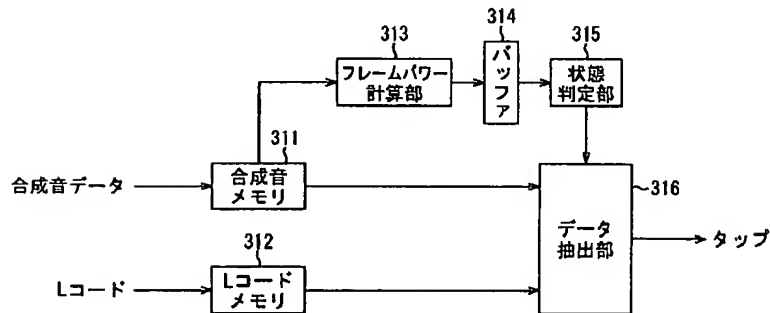


【図 11】



受信部 114

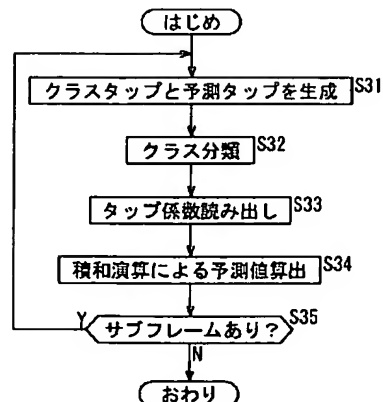
【図 13】



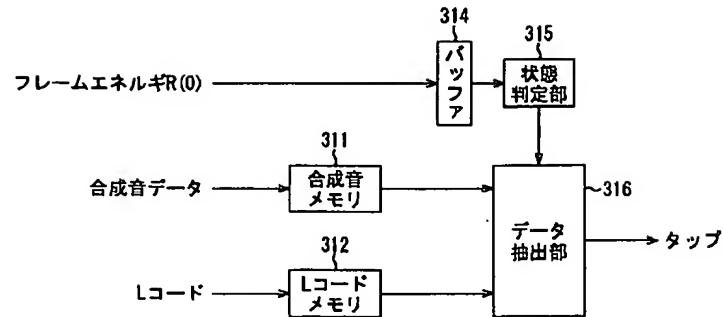
タップ生成部 301 (302)

【図 19】

図19

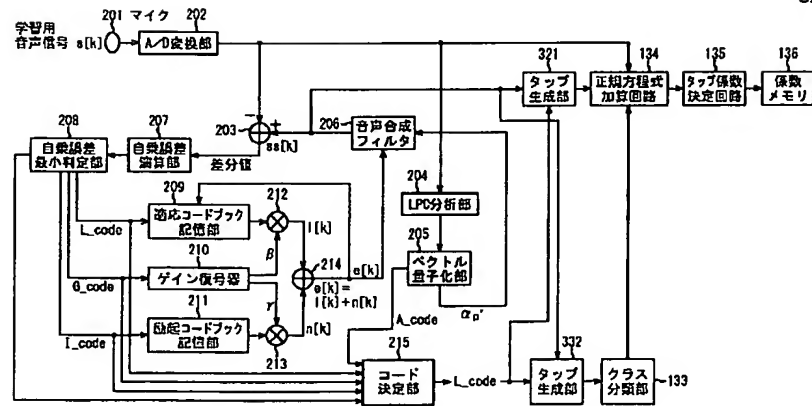


15



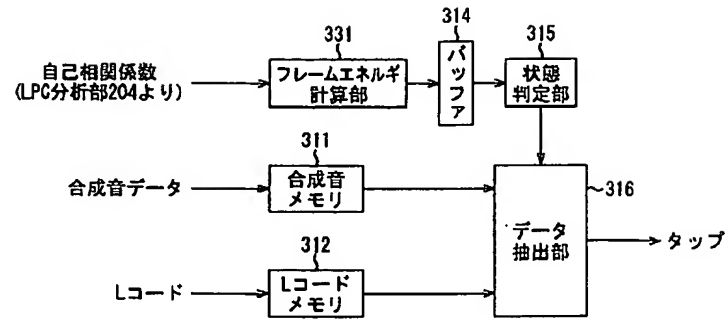
タップ生成部 301 (302)

图16



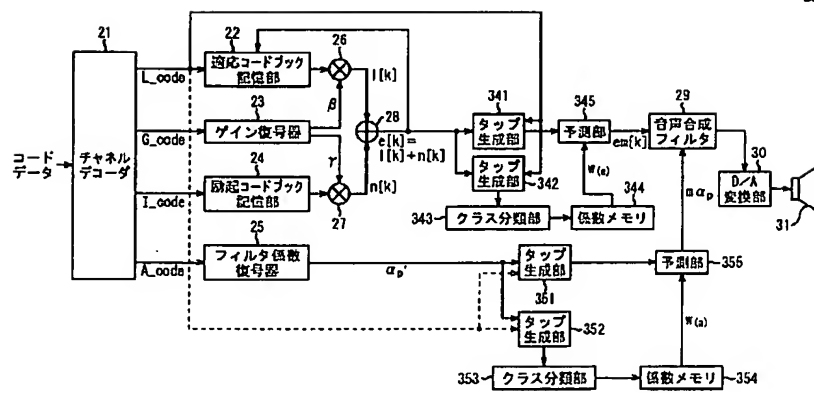
学習装置

【図 17】



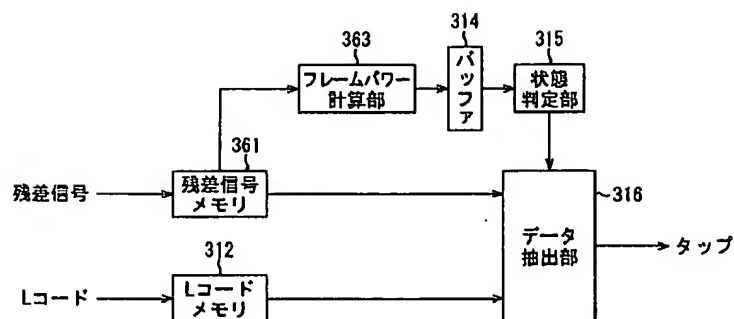
タップ生成部 321 (322)

【図 18】



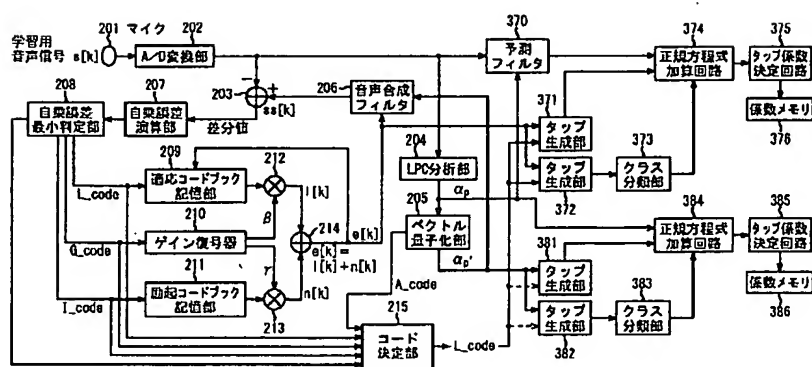
受信部 114

【図 20】



タップ生成部 341 (342)

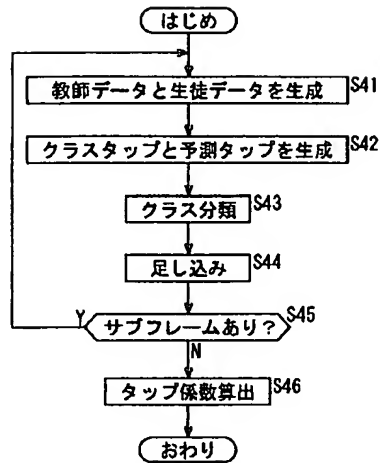
【图 2 1】



學習資源

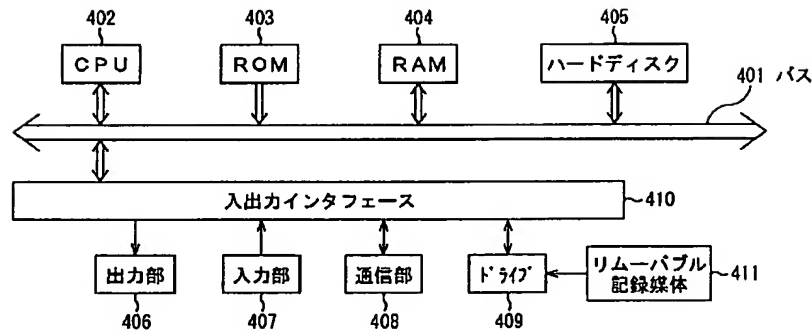
【図 22】

図22



【図 23】

図23



コンピュータ

フロントページの続き

(72)発明者 渡辺 勉
東京都品川区北品川 6 丁目 7 番35号 ソニ
ー株式会社内

(72)発明者 服部 正明
東京都品川区北品川 6 丁目 7 番35号 ソニ
ー株式会社内

Fターム(参考) 5D045 CA01
5J064 AA01 BB03 BC14 BC28 BD03